MAR 22 1922

Journal of Educational Research

OFFICIAL ORGAN OF THE NATIONAL ASSOCIATION OF DIRECTORS OF EDUCATIONAL RESEARCH

B. R. BUCKINGHAM, EDITOR

ASSOCIATE EDITORS

E. J. ASHBAUGH W. W. CHARTERS GEORGE D. STRAYER
GUY M. WHIPPLE S. A. COURTIS LEWIS M. TERMAN

VOL. V MARCH, 1922 No. 3

CONTENTS

A COMPARATIVE STUDY OF FOUR GROUP SCALES FOR THE PRIMARY GRADES ..... V. A. C. Henmon and Ruth Streitz 185

EDUCATIONAL MEASUREMENT AS A KEY TO INDIVIDUAL INSTRUCTION AND PROMOTIONS ..... Carleton W. Washburne 195

QUALITIES RELATED TO SUCCESS IN ELEMENTARY SCHOOL TEACHING ..... F. B. Knight 207

SOME USES FOR INTELLIGENCE TESTS ..... Samuel S. Brooks 217

THE VALIDITY OF THE WHIPPLE GROUP TEST IN THE FOURTH AND FIFTH GRADES ..... Helen Davis 239

EDITORIALS 245

REVIEWS AND ABSTRACTS 250

NEWS ITEMS AND COMMUNICATIONS 258

Entered as second-class matter at the Post Office at Bloomington, Illinois under the act of March 3, 1879.

mittee felt that it ought not to act upon such a matter without having the advice of the members of the association, who have had an opportunity to think over the matter very carefully before expressing an opinion. The publication of this communication at this time should give the members of the Association ample opportunity to consider the question from all angles and to be prepared to discuss intelligently the merits of the question at our meeting in Chicago. The executive committee will welcome communications from members of the Association relative to this question.

COMMITTEE TO APPRAISE STANDARD TESTS?

Do we want it? If it can be honest, scientific and practical—yes! If it is to promote some existing "bloc" of test promoters—No!

One of the working principles of such a committee should be that no test should be denounced until it has been carefully tried out.

How would such a committee be appointed? To whom would this committee be responsible? The education of the masses is intrusted to school superintendents and their teachers. How much will they have to say about this committee?

If the chief function of such a committee is to protect superintendents, then, "forget it," and give the time and effort to more useful pursuits. Superintendents are quite capable of judging for themselves whether a test is pedantic or practical.

On the other hand, if such a committee could be broad gauged, realizing that at the end of the first decade of tests and measurements there are no savants, and that so far only a beginning has been made, but that that beginning, with all its defects, has been a very important one,—with such a conception plus an honest desire to render service, such a committee would be a very desirable thing.

ERNEST C. WITHAM

Supt. Schools, Putnam, Conn.

[The above reply from Superintendent Witham is the only one that has come to us as a result of the invitation printed in the January number. (Page 83). We print it in full in order that our readers may get Superintendent Witham's reaction to the situation. Since this matter will come before the Association for discussion at one of the closed sessions, it is hoped that a large number of our members will be ready to express themselves on the proposition at that time.

We would like to call attention to one fact. No one is entitled to be considered an expert in any field because of the position which he holds. Expertness is an attribute derived from unusual knowledge and not from position held. The statements which we have received from large numbers of superintendents would indicate that a great many of them at least feel the need of guidance in the field of tests.

EDITOR.]

MAR 22 1922

Journal of Educational Research

OFFICIAL ORGAN OF THE NATIONAL ASSOCIATION OF DIRECTORS OF EDUCATIONAL RESEARCH

B. R. BUCKINGHAM, EDITOR

ASSOCIATE EDITORS

E. J. ASHBAUGH W. W. CHARTERS GEORGE D. STRAYER
GUY M. WHIPPLE S. A. COURTIS LEWIS M. TERMAN

VOL. V MARCH, 1922 No. 3

CONTENTS

A COMPARATIVE STUDY OF FOUR GROUP SCALES FOR THE PRIMARY GRADES ..... V. A. C. Henmon and Ruth Streitz 185

EDUCATIONAL MEASUREMENT AS A KEY TO INDIVIDUAL INSTRUCTION AND PROMOTIONS ..... Carleton W. Washburne 195

QUALITIES RELATED TO SUCCESS IN ELEMENTARY SCHOOL TEACHING ..... F. S. Knight 207

SOME USES FOR INTELLIGENCE TESTS ..... Samuel S. Brooks 217

THE VALIDITY OF THE WHIPPLE GROUP TEST IN THE FOURTH AND FIFTH GRADES ..... Helen Davis 219

EDITORIALS 245

REVIEWS AND ABSTRACTS 250

NEWS ITEMS AND COMMUNICATIONS 258

Entered as second-class matter at the Post Office at Bloomington, Illinois under the act of March 3, 1879.

Journal of Educational Research

Copyright 1922, by
Public School Publishing Co.

Official Organ of the National Association of Directors of Educational Research

Editor-in-Chief, **B. R. Buckingham**
Director of Bureau of Educational Research, Ohio State University, Columbus, Ohio

Associate Editors

E. J. ASHBAUGH, *Ex-Officio, Secretary of Research Association, Ohio State University.*

W. W. CHARTERS, *Carnegie Inst. Tech., Pittsburgh.*

S. A. COURTIS, *City Schools, Detroit, Mich.*

GEORGE D. STRAYER, *Columbia University, New York City.*

LEWIS M. TERMAN, *Leland Stanford Junior University, Stanford University, California*

GUY M. WHIPPLE, *College of Education, University of Michigan, Ann Arbor, Michigan.*

Contributing Editors

VIRGIL E. DICKSON, *Director, Bureau of Research and Guidance, Public Schools, Oakland and Berkeley, California.*

CHARLES FORDYCE, *Director of Bureau of Educational Measurements, University of Nebraska, Lincoln, Nebr.*

M. E. HAGGERTY, *Dean of the College of Education, University of Minnesota, Minneapolis.*

V. A. C. HENMON, *Director, School of Education, University of Wisconsin, Madison.*

ARTHUR W. KALLOM, *Assistant Director, Department of Educational Investigation and Measurement, Public Schools, Boston, Mass.*

F. J. KELLY, *Dean of University Administration, University of Kansas, Lawrence, Kans.*

ROBT. H. LANE, *Assistant Superintendent of Schools, Los Angeles, Cal.*

WILLIAM A. McCALL, *Assistant Professor of Education, Teachers' College, Columbia University.*

GEORGE W. MELCHER, *Director, Bureau of Research and Efficiency, Public Schools, Kansas City, Missouri.*

EUGENE A. NIFENECKER, *Assistant Director, Bureau of References, Research and Statistics, Public Schools, New York City.*

JOSEPH P. O'HERN, *Assistant Superintendent of Schools and Director of Efficiency Bureau, Public Schools, Rochester, New York.*

LEON O. SMITH, *Assistant Superintendent (in charge of Educational Research) Public Schools, Omaha, Nebraska.*

W. W. THEISEN, *Director, Division of Reference and Research, Public Schools, Cleveland, Ohio.*

CLIFFORD WOODY, *Director Bureau of Tests and Measurements, University of Michigan, Ann Arbor, Mich.*

The **JOURNAL OF EDUCATIONAL RESEARCH** is published at Bloomington, Illinois, monthly, except July and August by the Public School Publishing Co. The subscription price is $4.00; $4.20 in Canada or Mexico; and $4.50 in other foreign countries.

The editorial office is at Columbus, Ohio. Address all communications on editorial matters to B. R. Buckingham, Director of Bureau of Educational Research, The Ohio State University, Columbus, Ohio.

Subscriptions and all communications about reprints or other business matters should be addressed to the Public School Publishing Company at Bloomington, Illinois.

JOURNAL *of* EDUCATIONAL RESEARCH

Volume V MARCH, 1922 Number 3

A COMPARATIVE STUDY OF FOUR GROUP SCALES FOR THE PRIMARY GRADES

V. A. C. Henmon and Ruth Streitz
University of Wisconsin

The multiplication of group tests for measuring native intelligence proceeds apace. It is safe to say that more scales can be made in a month than can be adequately evaluated in a year. Departments of education and psychology are importuned for advice and counsel as to the validity, reliability, and relative merits of such scales as are in the market. The following study was undertaken to compare four group scales for the primary grades, viz., Pressey's Primer Scale,[1] Myers' Mental Measure,[2] Dearborn's Group Test, Series 1,[3] and Haggerty's Delta 1.[4] Several other tests, such as Kingsbury's and Cole's, have appeared since this investigation was begun.

A satisfactory group test for the first grade is difficult to construct because of the necessity for minimizing the appeal to abilities derived from schooling. A test for this grade, however, is particularly desirable, in view of the varying ages of entrance and the absence of other data, such as school marks or accumulated testimony of teachers, which might serve for purposes of classification and prognosis.

[1] Pressey, Luella W. "Mental Survey Tests—Primer Scale." Public School Publishing Company, Bloomington, Illinois.

[2] Myers, Garry C. "The Myers Mental Measure." Newson and Company, New York.

[3] Dearborn, Walter F. "Group Tests of Intelligence, Series I." J. B. Lippincott and Company.

[4] Haggerty, M. E. "Intelligence Examination, Delta I." World Book Company.

Data Secured

The four tests were given by one of the writers to one hundred pupils in two schools, fifty in two first grade classes and fifty in two second grade classes. The teachers in each of the four classes were asked to rate the pupils on the usual percentage scale in native ability. These ratings were secured rather than school marks which are notoriously inaccurate in the first two grades. Careful notes were made by the examiner of the reactions of pupils, ambiguities or difficulties in the tests, and ease of administration and scoring.

The distributions of the scores in the different tests with the medians, average deviations, coefficients of variability, and the probable errors of the medians, appear in Table I.

TABLE I. DISTRIBUTIONS OF SCORES IN FOUR GROUP SCALES

Score	Pressey		Myers		Dearborn		Haggerty	
	Grade I	Grade II	Grade I	Grade II	Grade I	Grade II	Grade I	Grade II
0-4	1							
5-9	2							
10-14	3		5					
15-19	1		6		1			
20-24	2		5	3	1			
25-29	1		7	3				
30-34	2		6	5		1		
35-39		1	5	9	3		1	1
40-44	2		12	6	4	1	2	
45-49	4	1	3	6	6		1	
50-54	5	1		4	10	3	1	1
55-59	6	3	1	6	14	2	3	1
60-64	7	6		3	6	7	3	1
65-69	6	10		4	3	18	4	
70-74	6	12				9	11	
75-79	2	12		1	2	7	5	3
80-84		3				1	4	1
85-89						1	5	3
90-94		1					2	5
95-99							2	9
100-104							2	10
105-109							3	8
110-114								3
115-119								4
Number	50	50	50	50	50	50	50	50
Median	56.5	71	32.5	44	54.5	67	72	99.5
A. D.	15.8	6.6	10.0	11.3	7.9	6.5	13.4	11.1
C. V.	.280	.093	.308	.240	.143	.097	.183	.125
P. E.	1.98	.77	1.44	1.41	.85	.64	1.65	1.13

Figures 1, 2, 3, and 4 show the distributions of Table I graphically.

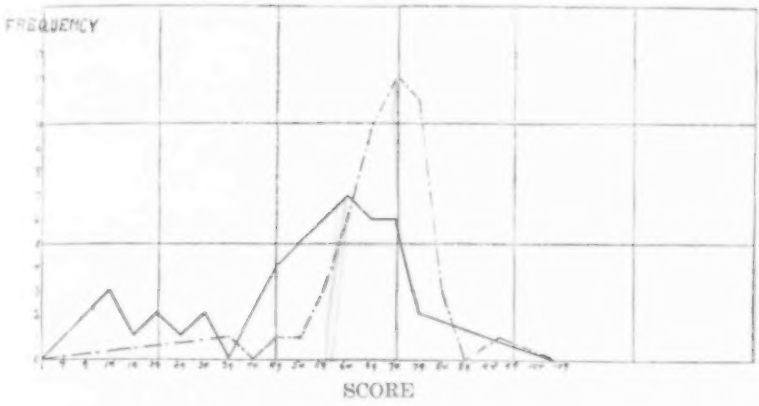

FIGURE 1. DISTRIBUTION OF SCORES ON THE PRESSEY TEST
(Full line represents first grade, dot-and-dash line represents second grade.)

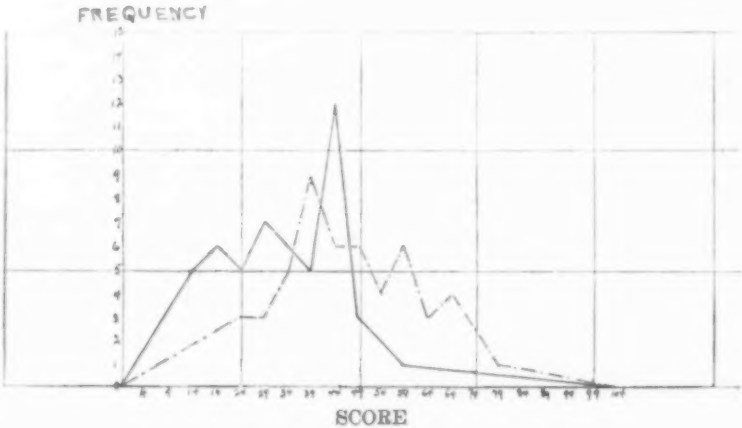

FIGURE 2. DISTRIBUTION OF SCORES ON THE MYERS TEST
(Full line represents first grade, dot-and-dash line represents second grade.)

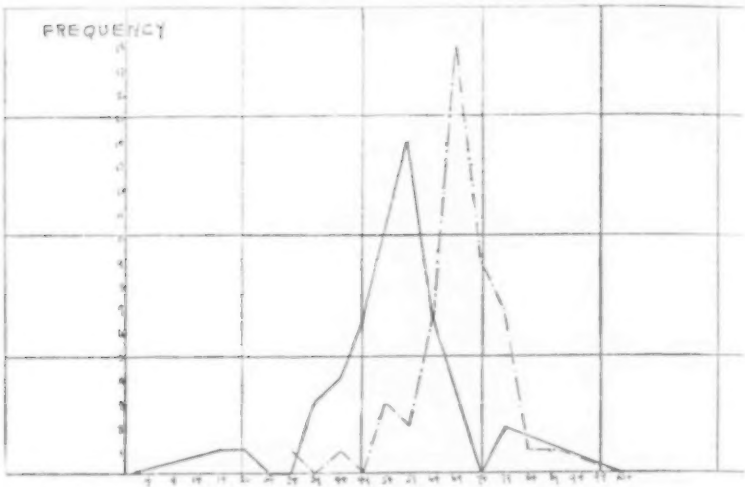

FIGURE 3. DISTRIBUTION OF SCORES ON THE DEARBORN TEST
(Full line represents first grade, dot-and-dash line represents second grade.)

FIGURE 4. DISTRIBUTION OF SCORES ON THE HAGGERTY TEST
(Full line represents first grade, dot-and-dash line represents second grade.)

With such results as these, what tests may be used to determine the relative merits of the different scales? In the absence of an accepted definition of what constitutes intelligence or of an absolute criterion against which the results with a scale may be checked, it is necessary to resort to indirect criteria. Various methods have been employed or proposed by Lowell,[5] Holley,[6] Mrs. Pressey,[7] Pressey,[8] and Kelley.[9]

Correlations with Teachers' Estimates

Other things being equal, that test will be most useful which correlates highest with some other criterion, such as school marks, teachers' ratings, scores in standard educational tests, or mental ages in individual tests. It has often been shown that there is a high correlation between performance in elementary school and in high school, and between high-school and college performance, as evidenced by school marks or teachers' estimates. If, then, scores in a group scale correlate highly with scholastic ratings or teachers' estimates, they would possess prognostic value and be useful for purposes of classification.

A common method, where the data are available, has been to measure the validity of a group scale by its agreement with the results of the Binet-Simon examination or other test series of supposedly great validity. This involves the assumption that the individual examination is a more reliable measure of intellect than the group scale or is a closer approximation to an absolute criterion. In view of the fact that the group scales employ somewhat the same material adapted to group use, testing the validity

[5] Lowell, Frances. "A group intelligence scale for primary grades." *Journal of Applied Psychology*, 3: 215-47, September, 1919.

[6] Holley, C. E. *Mental tests for school use.* University of Illinois Bulletin, No. 28, March, 1920, pp. 1-91.

[7] Pressey, L. W. "A group scale of intelligence for use in the first three grades: its validity and reliability." *Journal of Educational Research*, 1: 285-94, April, 1920.

[8] Pressey, S. L. "Suggestions looking toward a fundamental revision of current statistical procedure as applied to tests." *Psychological Review*, 27: 466-72, November, 1920.

[9] Kelley, T. L. "The reliability of tests." *Journal of Educational Research*, 3: 370-9, May, 1921.

of one scale by its agreement with another is not particularly satisfying.

The correlations with the teachers' estimates are given in Table II.

TABLE II. CORRELATIONS WITH TEACHERS' ESTIMATES

Test	First Grade		Second Grade		Average
	School A	School B	School A	School B	
Pressey	0.78 ± 0.05	0.74 ± 0.06	0.16 ± 0.13	0.60 ± 0.08	0.57
Myers	0.70 ± 0.07	0.24 ± 0.12	0.68 ± 0.07	0.43 ± 0.12	0.51
Dearborn	0.61 ± 0.08	0.71 ± 0.07	0.49 ± 0.10	0.71 ± 0.07	0.63
Haggerty	0.46 ± 0.11	0.58 ± 0.08	0.48 ± 0.10	0.60 ± 0.08	0.53

The correlations with the Pressey test are considerably higher than those reported by Holley, which were 0.42 ± 0.04 for the first grade and 0.30 ± 0.05 for the second grade. They are higher, except in the second grade, than those reported by Mrs. Pressey, namely 0.66 for 148 primary-grade children and 0.62 for 57 six-year-olds.

Aside from the Pressey test in the second grade of School A and the Myers test in the first grade of School B, the correlations are high and do not differ greatly. The Pressey test gives the best results with the first grade but does not seem so well adapted for the second grade. The Dearborn test, which has limitations as to cost, difficulty of administration and scoring, and lack of interest by children, nevertheless gives the best results on the average for both grades.

In view of the differences in the results, it became a matter of interest to determine the intercorrelations between the tests. They are given for each school and grade in Table III. In each case these four coefficients are averaged and entered in the lower left-hand half of the table.

The rather low intercorrelations indicate that the tests measure different phases of ability. This suggested the possibility of combining two or more tests into a team with the prospect of raising the coefficients with teachers' estimates. Combining all four tests, the scores being weighted so as to give approximate equality to each test, gave the following results for the first

TABLE III. INTERCORRELATIONS BETWEEN THE DIFFERENT TESTS

	PRESSEY	MYERS	DEARBORN	HAGGERTY
Pressey		0.69 0.37 0.44 0.43	0.50 0.77 0.29 0.75	0.30 0.69 0.24 0.60
Myers	0.48		0.62 0.53 0.65 0.53	0.33 0.21 0.28 0.69
Dearborn	0.58	0.58		0.82 0.67 0.27 0.72
Haggerty	0.46	0.38	0.62	

grade: School A, 0.79; School B, 0.65. The corresponding coefficients for the second grade were: School A, 0.71; School B, 0.69. The average of these four coefficients is 0.71.

It is noteworthy that this team of tests did not give as satisfactory a result in the first grade as the Pressey test did alone. In the second grade the Dearborn test in one school gave a better correlation than the team. The average result, however, for the four classes is very distinctly better than with any single test. It is better than any two of the tests combined, as is shown in Table IV. The average results, however, are distinctly better with a pair of tests than any single test.

TABLE IV. CORRELATIONS WITH TEACHERS' ESTIMATES FOR VARIOUS PAIRS OF TESTS

PAIRS OF TESTS	FIRST GRADE		SECOND GRADE		AVERAGE
	SCHOOL A	SCHOOL B	SCHOOL A	SCHOOL B	
Pressey-Myers	90	53	63	60	66
Pressey-Dearborn	81	75	86	71	66
Pressey-Haggerty	71	88	43	63	66
Myers-Dearborn	83	49	66	71	67
Myers-Haggerty	63	50	71	55	60
Dearborn-Haggerty	60	69	58	63	62

It is difficult to draw definite conclusions from the above data. Success in school work, on which teachers' estimates of ability are likely to rest, depends on many other factors in addition to intelligence so that checking the value of intelligence scores by their agreement with teachers' estimates or scholastic ratings is at best uncertain. Moreover, the method requires the assumption that

the teachers' estimates are a perfect criterion, which they are not. The test which agrees best with teachers' judgments is not, then, necessarily the best measure of intelligence. Other criteria of validity must therefore be used.

Discriminative Capacity

Other things being equal, that test is most useful which segregates most clearly and unambiguously different age or grade groups. The larger the difference between the averages or medians of six year olds and seven year olds or between first-grade and second-grade children, the better the test for classificatory purposes. More important than the difference between the central tendencies of two age or grade groups is the variability in each group. The smaller the coefficient of variation or the smaller the percent of overlapping between successive years, the greater the discriminative capacity and usefulness of the test. For practical purposes in these respects a comparison of grade groups is more useful than a comparison of age groups.

An inspection of the curves in Figures 1 to 4 shows the overlapping between the grades and indicates that from this point of view the Myers test is distinctly inferior to the other three. The facts are shown better by the percent of first-grade children who reach or exceed the median of the second grade—Pressey 12 percent, Myers 14 percent, Dearborn 6 percent, and Haggerty 10 percent—and by the percent of second-grade children who fail to reach the median of the first grade—Pressey 6 percent, Myers 20 percent, Dearborn 10 percent, and Haggerty 8 percent. The averages of these percents as the best indices of overlapping are as follows: Pressey 9 percent, Myers 17 percent, Dearborn 8 percent, and Haggerty 9 percent.

Another method of measuring discriminative capacity (and this is the method which Holley used) is to determine for each scale the differences between the grade medians and the probable errors of the differences, and to render these differences comparable by dividing the probable errors of the differences by the differences themselves. This gives a coefficient of variability.

The smaller this coefficient the more discriminating the test. The facts for the four tests, being the differences between the grade medians, the probable errors of these differences, and the resulting coefficients of variation, are given in Table V.

TABLE V. DATA ON THE DISCRIMINATIVE CAPACITY OF FOUR TESTS

Test	Difference Between Medians	P.E. of Each Difference	Coefficient of Variation
Pressey	14.5	2.12	0.146
Myers	11.5	2.01	0.175
Dearborn	12.5	1.06	0.084
Haggerty	27.5	2.00	0.073

These data confirm the conclusion, which might be reached from an inspection of Figures 1 to 4, that the Dearborn and Haggerty Tests segregate the first-grade children from the second-grade children considerably better than the other two tests.

Character of the Distribution Curve

The validity of a scale has been tested by the conformity of the scores to the normal distribution curve. A group scale to have discriminative capacity must be of such difficulty as to register individual differences clearly for any year or grade. A markedly skewed distribution with many failures or very low scores or with many perfect or very high scores would be useless. While there is nothing sacrosanct about the normal distribution curve, other things being equal, the test that approximates most closely the normal distribution is, perhaps, the best.

Pressey attacks vigorously the notion that conformity to the normal curve is evidence of the validity of a scale for practical purposes of classification. The validity of a scale as a measure of general intelligence is a different matter from the value of the scale for practical purposes of diagnosis and classification. For practical purposes a scale would be most useful if it "sets off the potential failures at the bottom and the assured successes at the top and spreads out the questionable cases in between." The dis-

tribution in fact, to be desired, is bimodal, not normal. Kelley similarly says conformity to the normal distribution is utterly inadequate as a test of reliability.

There is much to be said for measuring the value of a scale by its success in accomplishing some specific purpose rather than by its conformity to some theoretical standard. It is doubtful, however, whether a test of general intelligence could be devised that would give a bimodal distribution. Certainly from all we know of the distribution of mental traits it would be unique.

The number of cases in this study is not sufficiently large to compare the distribution curves satisfactorily but an inspection of Figures 1 to 4 and Table I shows fairly close conformity to the normal curve except in the case of Myers test.

Administration and Scoring

It is clear from the above data that there are no very striking differences between three of the scales. None of the three is a perfect measuring instrument but each gives information that is well worth while securing. In view of the fact that there is no outstanding superiority of one test over the other two, the factors of each of administration, time of scoring, and reactions of pupils will have weight with teachers and superintendents. According to the examiner's reports, the Pressey test ranks high in these respects since it can be given in twenty-five minutes and scored in three minutes. The Haggerty test also ranks high as it can be given in thirty minutes and scored in five minutes. Both hold the attention of children reasonably well. This is noticeably true of Haggerty's test. The Dearborn test, which ranked high in the tests applied, is long and requires that it be given in three sections. It takes about fifteen minutes to score. In administration, it requires preparation, care, and skill on the part of the examiner, and is fatiguing and boresome to many of the children, especially in the first grade.

EDUCATIONAL MEASUREMENT AS A KEY TO INDIVIDUAL INSTRUCTION AND PROMOTIONS

CARLETON W. WASHBURNE
Superintendent of Schools, Winnetka, Illinois

Time has been the one measurable factor in education until recently. Consequently, we have laid out our educational system in terms of time units. These have been constant, while the achievement of the pupils within these time units has varied according to individual ability. But with the development of the achievement test movement, it is becoming possible to reverse this order. We may now make units of achievement the constant factor, varying the time to fit the individual capacities of the children.

Educational measurements as now known to most of us, however, are not yet sufficiently developed to enable us to make the complete change in our schools which is implied by variable time units and constant achievement units. Yet the movement toward this change is unmistakable, and in the schools of Winnetka we have been experimenting for the past three years on the possibility of making the change complete. Whatever the crudities of the detail in our procedure, we have demonstrated the practicability of a shift from time units to achievement units. We have for nearly two years succeeded in making all of our promotions individual and on the basis of achievement rather than of time.

The firm hold of the time unit upon our educational system dates back a long way, but has grown especially strong with the development of graded city schools. As a consequence of making time the constant factor, rapid pupils either acquire habits of comparative idleness and work below their ability, or are pushed ahead a grade, losing much valuable intermediate instruction; slow pupils are either advanced with inadequate foundation or are retarded abnormally by grade repetition. A scheme which does not take advantage of the natural alertness of the bright

pupils and which clogs the system with repeaters is financially and educationally wasteful. The faults which result from the class lockstep are too well known to need discussion. We all earnestly wish to abandon this system, but we have not heretofore found a safe method of doing so.

Before educational measurements were developed several attempts were made to individualize instruction and promotions. The best known of these were the Pueblo, Colorado plan, developed by Preston Search about thirty years ago, and the San Francisco State Normal School plan developed by Frederic Burk about ten years ago. Neither of these made time the constant factor; nor had they developed the technique of making achievement units the constant. Instead, they had work units. Each pupil had a certain amount of material to go through—a number of pages to read, a number of exercises to fill out, or a number of problems to work. Tests at the end of this work were of a sampling character, neither complete nor diagnostic. The difficulty of keeping up with the practice work and checking the assignments of the children was such that these systems failed to spread as rapidly as their merit deserved. The San Francisco Normal School plan was far in advance of the Pueblo plan because the material was well organized and specially prepared for the children. But until about four years ago this plan was such that small classes were strictly necessary, in order to make it administratively feasible.

Rural schools have always been individual in the same sense that the Pueblo plan was individual. Pupils could advance at their own rate, but each child went through the same work reciting on it orally to the teacher, or being tested by the random sampling process. The same is true of the high school plan of Dalton, Massachusetts.

With the ability to measure educational achievement accurately a new type of individual instruction and promotions has come into existence. The San Francisco State Normal School was one of the first institutions to undertake a systematic development of the measuring movement towards an administrative solu-

tion of the class lockstep problem. The methods which I shall describe below as the Winnetka plan had their inception at the San Francisco State Normal School under Frederic Burk. Much of what Winnetka is doing is closely paralleling the recent developments in this school.

In undertaking to introduce individual promotions and instruction in the Winnetka Public Schools in 1919 we were confronted by the inadequacy for our purpose of most of the standard tests on the market. They had been constructed largely from the standpoint of supervision. "Measure the efficiency of the entire school, not the individual ability of the few," was, and still is, the published motto of the author of some of the best tests. We wanted to reverse this procedure, and to measure the individual, not the school. For this purpose we found that we had to develop tests and measurements further than they had yet been developed. It became necessary to make each test cover completely the subject that was being tested, in such a way that a diagnosis of the difficulties of the individual pupil was easy for any teacher. Moreover, it was also necessary to prepare practice material which would correspond with each topic or process contained in the test so that a pupil might practice on his own specific weakness. It was desirable that this material should be as nearly as possible self-instructive, so that pupils could use it with the minimum of help from the teacher.

Before either tests or practice materials could be prepared it was necessary to establish our units of achievement. We call these units, "goals." Wherever possible we established them upon the basis of research already done. In many cases, however, where research was lacking, we used our best judgment in establishing temporary goals, and proceeded to undertake investigations which would gradually replace these temporary standards by others scientifically established. This is a long, slow process and is still in its early stages. Space in this article will not permit a discussion of these investigations. They will, however, be published from time to time as they are completed.

Before discussing the place of these achievement units in the Winnetka plan, it should be made clear that they represent only a part of the curriculum. An equally important part consists of what we call "social work." The social part of our work is not measured. It consists of that range of subject matter to which we wish to expose our children, but which they need not master, and of opportunities to enter into group undertakings which will develop the children's ability to co-operate with one another. The social part of our work is not "socialized recitation," because there is no recitation in it—children are not being tested on what they have studied. They are not re-citing. They are neither marked nor promoted on this basis. Social work consists of class discussions, group projects, dramatizations, special reports, group games on the playground, assemblies, group singing—all activities in which the group acts as a whole and in which each child is contributing his share to the group. The social work in the Winnetka Schools occupies from one-third to one-half of each school day, and is as important a part of the curriculum as the individual work. We have more opportunity for children to express themselves and to develop social attributes than systems which use the class recitation, because the individual work on minimum essentials economizes the children's time. The social work is also freed from the incubus of discovering whether or not children have studied their lessons.

The achievement unit takes the place of the time unit in the individual work on minimum essentials. Children's marks and promotions are based entirely upon individual work. There are no recitations; there is no grade repetition; there are no failures; there is no skipping. No child is held back to a slower rate of progress than is natural to him; none is forced forward too rapidly for thorough work. Each "goal" must be achieved by each pupil before he can go on to the next goal.

Instead of reciting the children practice and prepare themselves for achievement tests. The teacher instead of listening to recitations passes about among the children, helping them in the preparation of their materials, developing new work with small

informal, temporary groups of children who happen to be ready for the same instruction at the same time, and helping each child to work independently toward a definite end.

The plan is made administratively possible by the achievement tests corresponding to each goal in each subject and by practice materials which the children themselves can correct and which enable them to get ready for these achievement tests.

It has been necessary to prepare a complete series of achievement tests in arithmetic and language. None of the published tests in these two subjects were sufficiently detailed and diagnostic for individual instruction. The nearest approaches are the Courtis and Woody so-called diagnostic tests in arithmetic, and the Pressey achievement tests in language. Courtis or Woody may show us that a pupil is weak in addition or even that he is weak in column addition. Neither of them, however, will show us what particular combinations he is missing, so that he may correct his individual defect. They may show us that a child is weak in simple multiplication, but they will not show us whether the difficulty is in carrying or in lack of knowledge of the multiplication facts. Neither will they show us in which particular facts or carrying combinations the pupil is making his failures. The Pressey language tests do not as yet completely cover the minimum essentials of punctuation and capitalization, nor are the answers so keyed that it is possible for the pupil to practice on the particular error which he has made. These statements are unfair if taken as criticisms of the Pressey, Woody, or Courtis tests. Those tests are for the most part avowedly to test the school or the group, rather than to diagnose the specific difficulty of each individual. For our purposes, however, and for any school which wishes to make units of achievement the constant factor in place of units of time, it is necessary to have specifically diagnostic tests for each achievement unit.

The preparation of the tests we are using in Winnetka was begun by the faculty of the San Francisco State Normal School while the writer was director of tests and measurements at that institution. It has been continued by the Winnetka teachers dur-

ing the past three years. The criteria to which each test must conform are: (1) it must cover completely every detail of the unit tested—*e.g.*, an addition test must include every combination, a short division test every type of difficulty, etc.; (2) the test must diagnose readily the exact weakness of each pupil in so far as this weakness is a lack of knowledge or ability; (3) the test must be strictly objective and easily corrected; (4) the answers must be so keyed that the pupil can readily turn to corresponding practice material and make up his deficiencies.

The preparation of practice materials to correspond with these tests has been a very considerable task. It has been done co-operatively by groups of teachers, their work being mimeographed and distributed to the pupils. The practice materials have had to conform to the following criteria: (1) they must be readily intelligible to the children themselves, self-instructive so far as possible; (2) they must take up just one new thing at a time, providing ample practice in that new element before the next element is discussed; (3) they must be self-corrective, *i.e.*, they must be provided with answer sheets or other devices so that the child can correct all of his own practice work; (4) they must lead up to practice tests which shall be alternate forms of the regular test, so that a child can determine for himself whether or not he is ready for a regular test; (5) they must be arranged to correspond with the arrangement of the tests, so that a pupil may readily find the practice exercises corresponding with any error made in a test.

While we have prepared a series of tests and practice books for all grades in arithmetic and language (including grammar), we have not had to do so in reading and spelling. The Burgess Picture Scale test in silent reading makes it possible to measure the children's achievements in this subject with considerable accuracy, as does the Gray Oral Reading Test in oral reading. The Anderson Spelling List as arranged in the Iowa Scale makes the measurement of achievement in spelling definite and simple.

In the first two grades, however, it has been necessary to prepare special devices and tests for reading. Descriptions of these

will be published as they are perfected. Above the second grade a pupil reads fifteen books of average length and difficulty for each grade. He then takes the Burgess and Gray Reading tests. If his achievement is equal to the standard achievement of the next grade above, he is promoted to that grade. We are working on exercises to increase comprehension where this is low, and to increase speed where the child reads too slowly. We have undertaken a rather elaborate investigation of books for children, hoping to standardize them according to difficulty and merit.

Our procedure in spelling consists in dictating all the words for the grade at the beginning of the year, twice through. Those words which a pupil misses are checked in his individual speller and become his course of study for the year. In the back of his speller are blanks for words misspelled in his written work which also become part of his course of study. Pupils dictate to each other in pairs and correct their own daily work. Once a week they dictate the week's words to each other, giving these papers to the teacher for correction. Words spelled correctly on a weekly review test are not again studied or tested until the beginning of the next semester. In September and February there are semester review tests in which pupils dictate to their partners all the checked words in the partner's speller, the papers being corrected by the teacher. Words spelled correctly on these semester review tests do not have to be studied and will not be retested until the next semester review test. A word spelled correctly for four semester review tests is permanently removed from the pupil's course of study in spelling.

The effect of concentrating the pupil's attention on the words which he has misspelled has been to increase the general level of the children's spelling ability in Winnetka, until our grade medians are now from one to two years above the standards of the Iowa Scale. Likewise our individual method of teaching reading has resulted in grade medians from one to two years above the standards for the tests used.

History and geography are social subjects to a very considerable extent. There is, however, a certain essential body of

facts, which every child should know and which therefore should be individualized. We are undertaking, with the help of the Commonwealth Fund in New York, an extensive investigation as to what these facts are; and we hope ultimately to develop a technique for handling them which will not deprive history and geography of their social value nor result in lack of thoroughness in respect to the body of essential facts in each of these subjects.

Writing is individualized by means of a device for measuring the uniformity of slant, height of letters, position on line, non-interference of loop letters below the line with letters on the line below, speed, and legibility. Each of these is measured objectively by the pupils themselves as well as by the teachers, and pupils keep definite records of their improvement from day to day in the particular elements in which they are weakest. A transparent celluloid card four by six inches ruled the long way with twelve lines to the inch is used for measuring uniformity of slant, and height of small and tall letters.

Special subjects are also expressed in terms of definite units of achievement. Cooking, sewing, and general science all lend themselves readily to this system.

To make the operation of this individual work concrete, let us consider for a moment what you would find if you were to step into a fifth-grade room in Winnetka some morning when most of the children were studying language. You would find one child working on the latter part of his fourth-grade language book. Another might be beginning his sixth-grade language work. Most of the children would be scattered through different parts of the fifth-grade work. If you were to watch one of these you would see that he has before him a mimeographed practice book. He is seen to read first the development material on, say, "Commas in Series." He then practices writing sentences containing series, placing commas in their proper places, referring to the answers in the back of his book, and correcting his own work. After this specific practice in commas in series he takes a practice test. This is a part of his practice book and consists of a paragraph with all capitals and punctuation omitted;

it requires knowledge of every element up to and including commas in series. He copies this paragraph in the correct form, and compares it with the correct copy in the answer part of his book. This correct copy is keyed to show in what elements the child is weak. A sample test and keyed answer sheet is reproduced herewith.

FIFTH GRADE LANGUAGE BOOK—p. 8.

Cumulative Practice Test—16-A

a walk in the woods

one day marys mother told her she might ask francess mother to let her go nutting the girls mothers packed a lunch for them and off they started it was a sunny october saturday the girls lived in winnetka illinois dont you think the childrens holiday was a happy one they met mr jones and he walked home with them. The girls brought nuts bright leaves and fall flowers.

Cumulative Practice Test—16-A—Answer Sheet

A Walk in the Woods.
$\overline{7}$ $\overline{7}$ $\overline{7}$

One day Mary's mother told her she might ask Frances's mother
$\overline{4}$ $\overline{6}$ $\overline{8}$ $\overline{6}$ $\overline{8}$

to let her go nutting. The girls' mothers packed a lunch for them
$\overline{1}$ $\overline{4}$ $\overline{14}$

and off they started. It was a sunny October Saturday. The girls
$\overline{1}$ $\overline{4}$ $\overline{12}$ $\overline{11}$ $\overline{1}$ $\overline{4}$

lived in Winnetka, Illinois. Don't you think the children's holiday
$\overline{5}$ $\overline{13}$ $\overline{5}$ $\overline{1}$ $\overline{4}$ $\overline{9}$ $\overline{15}$

was a happy one? They met Mr. Jones and he walked home with
$\overline{3}$ $\overline{4}$ $\overline{10}$ $\overline{2}$ $\overline{6}$

them. The girls brought nuts, bright leaves and fall flowers.
$\overline{16}$

Analysis:

Goal 1....4 times.
Goal 2....1 time.
Goal 3....1 time.
Goal 4....6 times.
Goal 5....2 times.
Goal 6....2 times.
Goal 7....1 time.
Goal 8....2 times.
Goal 9....1 time.
Goal 10....1 time.
Goal 11....1 time.
Goal 12....1 time.
Goal 13....1 time.
Goal 14....1 time.
Goal 15....1 time.
Goal 16....1 time.

If he finds that he has forgotten his apostrophe for contractions (9), he turns back to section 9 of his practice book and practices on that and on each other error that he has made, and then asks the teacher for a test. This test is another form of

the practice test. After taking the test, he hands it to the teacher for correction. The next morning she returns it to him with all the errors marked. He turns back to his practice book and practices specifically on each point missed, and then gives himself a second form of practice test which he finds in his practice book. After correcting this, he asks for a second form of the regular test. If he gets it all right, he is marked O.K. in this topic in his "goal book." If he still makes errors, he repeats the first procedure. After receiving an O.K. in commas in series, he proceeds to the next goal, which happens to be "comma in address."

We have spoken of the "goal book." This is an abbreviated course of study stated in terms of specific achievement and placed in the hands of *pupils*.

Moreover, it is sent home monthly in lieu of a report card. Parents therefore as well as pupils know exactly what has to be accomplished and how much has been accomplished. Opposite each page of goals is a page of explanation to the parents.

The language page of the fifth-grade goal book looks like this:

FIFTH GRADE LANGUAGE GOALS

1. Period—end of sentence	13. Comma—city, state .....
2. Period—abbreviation	14. Apostrophe—"Boys'" .....
3. Question Mark	15. Apostrophe—"Children's" .....
4. Capital—begin sentence	16. Comma in series .....
5. Capital—place	17. Comma in address .....
6. Capital—person	18. Comma—yes, no, Oh, ...
7. Capital—title	19. Simple quotation .....
8. Apostrophe—"John's"	20. Exclamation point .....
9. Apostrophe—contractions	21. Capital for Diety .....
10. Capital—Mr., Captain Smith, etc.	22. Paragraphs .....
11. Capital—day of week	23. Use of Dictionary .....
12. Capital—months	24. Review .....

Course begun.....192....

Promoted to Grade VI Language.....192....

.....

Teacher

The arithmetic page of the fifth-grade goal book is shown below. "Speed" means the number of examples of standard difficulty worked in three minutes; "accuracy" is the percent correct.

FIFTH-GRADE ARITHMETIC GOALS

Addition review—		Multiplication of fractions—	
Speed 5, Accuracy 90%...	.....	Speed 4, Accuracy 90%...	.....
Subtrac. review—		Division of fractions—	.....
Speed 9, Accuracy 90%...	.....	Speed 4, Accuracy 90%...	.....
Simple Multip. review—		Decimals—values, addn. &	
Speed 3, Accuracy 90%...	.....	sub.—	
Compound Multip. review—		Speed 4, Accuracy 100%..	.....
Speed 5, Accuracy 90%...	.....	Decimals, frac. to dec., &	
Short Division review—		mult.—	
Speed 10, Accuracy 90%..	.....	Speed 4, Accuracy 100%..	.....
Long Division review—		Decimals—division—	
Speed 2, Accuracy 90%...	.....	Speed 4, Accuracy 100%..	.....
Addition of fractions—		Decimals—short cuts—	
Speed 4, Accuracy 90%...	.....	Speed 10, Accuracy 100%..	.....
Subtraction of fractions—		Problem Solving	.....
Speed 4, Accuracy 90%...	.....	Measurements	.....

Course begun.....192....

Promoted to Grade VI Arithmetic.....192....

.....

Teacher

The grouping of children in rooms is in accordance with their general advancement. There is nothing to prevent a child in one room from working on fifth-grade language and sixth-grade arithmetic at the same time. A child does not change rooms each time he receives a promotion. Room groupings are usually changed once a year, but no general rule is necessary. The individual work makes it possible to fit each child with that group to which he seems to belong. Very precocious children are usually placed in a room with children of an age which corresponds with the average of the chronological and mental ages. Most attempts to give precocious children the type of school work which they need brings them into social contact with classmates who differ markedly from them in every quality except intellectuality. The flexibility of the system we are now using at Winnetka obviates this difficulty. Other things being equal, however, children are placed in a room with other children who are doing approximately the same grade of work in most subjects. The most important factor in grouping children is their social work in which they have to co-operate with one another.

One of the outstanding features of the Winnetka plan of individual work is that each step toward its establishment is a

step worth while in itself, perfectly safe for any superintendent to take in any school system. The first of these steps is the establishment of definite goals or subject-matter units. This is a natural outgrowth of the educational measurements movement. The second step is also a development of the same movement—the preparation of tests which will completely cover each subject-matter unit and diagnose the difficulties of each individual child. The third step is the preparation of self-corrective practice materials which will at once prepare for these tests and enable a child to make up the deficiencies shown by the tests. When one has taken these three steps there is no difficulty whatever in placing an entire school system on an individual basis. Achievement units become the constant factor almost automatically. For after all, the time unit is the arbitrary one, the achievement unit the natural one. When achievement replaces time as the constant factor in our school systems, we can promote children individually by subjects and fit our public schools to the needs of the individual child.

QUALITIES RELATED TO SUCCESS IN ELEMENTARY SCHOOL TEACHING

F. B. KNIGHT
State University of Iowa

What facts concerning a candidate for a teaching position are of prognostic value? Of a hundred graduates of a normal college quite probably some will make excellent teachers, a larger number will do well, and a few will fail. By what system of interviewing can a superintendent increase his chances of picking more successful teachers and fewer failures than pure chance would account for? What qualities possessed by a candidate and ascertainable by a prospective employer are correlated highly enough with teaching success to be worth considering in a sound selective technique?

The study here reported was undertaken with such questions in mind, the purpose being to contribute statistically dependable facts as to teacher selection. The data were obtained from three school systems in Massachusetts which we shall call A, B, and C. Town A is a community of some 12,000 inhabitants, close to Boston. Town B is a community of about 10,000. The manufacturing interests of these two towns make them above the average "going" communities. Town C is a community of about 8,000, of great historic fame but with interests centering rather definitely in the past. The superintendents of Towns B and C were young men, university trained, with genuine professional interests. The writer was superintendent of Town A. The conditions for examining and testing the teachers were particularly fortunate in that the school boards were not only willing for the superintendents to make such studies but desired it. The investigation was under the direction of Professor E. L. Thorndike of Teachers College, New York.

The basic technique was to correlate ascertainable facts concerning the teachers with the degree of success they were obtaining in actual classroom work. Thus the investigation divided itself into two parts: first, to determine the distribution of teaching ability for each teaching staff; second, to correlate against this distribution the amounts of certain traits or qualities which

the individual teachers possessed. Later the partial correlation formula was used to determine the relation between success in teaching and each trait when other traits were rendered constant.

The Determination of Varying Amounts of Teaching Success

It is clear that we do not need for our purposes measures of teaching skill in absolute amounts. What we want and what we can get is the amount of difference between individual teachers of a staff. If we assign to the worst teacher in the elementary school of Town A a teaching ability of 1, and build up a consistent scale of merit so that the average teacher has, say, an ability of 15 and the best teacher an ability of 27, it is not important whether the arbitrarily chosen 1 represents a very high degree of teaching success compared with all the teaching in the United States or a small amount of skill. What we must have is amounts of difference, defensibly obtained, between members of the group.

The estimation of candidates has been far less serviceable both in industry and in education because employers have been confused in the matter of "amounts of qualities." In physical measurements we have units entirely objective and independent. Thus a teacher may weigh 130 pounds or be 5 feet 2 inches tall, and we know in this case exactly what we mean. But when we say a teacher is "good in explanation," "strong in discipline," or "weak in community interest," we are not so sure of what we mean. For instance, does "strong in discipline" mean a number of units above an absolute zero? As a matter of fact qualities, for which we do not have very objective units of measure, are consciously or unconsciously measured with reference to these qualities as people possess them. Strong in discipline means better in discipline than most teachers, I have known—above the average or some such thing. We have to measure many vital traits in this way. We can do it blindly and get comparisons which are very inaccurate and of little use; or we can do it accurately and express differences in mathematical terms so that these differences can be used for finer studies based on correlation. For in correlation it is useful to know not only differences

expressed in verbal adjectives but amounts of differences expressed in mathematical nouns, such as 4 units better in a spread from 1 to 30.

In this study amounts of differences were determined for six teaching groups in respect to general teaching ability. The six groups were the high-school staffs and the elementary-school staffs of the three school systems.

The final ratings for teaching success in each group were the result of the following procedure. We shall use the elementary-school group of Town A as an illustration.

Mutual ratings.—Every teacher in the group rated every other teacher for such qualities as (a) general teaching ability, (b) skill in discipline, (c) excellence of professional preparation, (d) ability to handle situations.

Upon prepared sheets containing the names of all the teachers together with columns for the several qualities each teacher rated every other teacher on each trait. The worst teacher in respect to a trait was to be marked 1, the next to be worst 2, and so on. The best teacher was to be rated 56 (there were 56 teachers in the group). Where no difference between two teachers could be discerned, the same mark for both was allowable. If a teacher was unknown to the rater, no mark was to be given. The rater was instructed to leave her rating sheet unsigned, and absolute freedom from any possible embarrassment was guaranteed.

The spread of ratings received was excellent. Only one teacher failed to cooperate. This teacher rated every teacher 1 in all qualities. The average spread was from 1 to 37. This was easily sufficient for building up an order of merit, quantitative in units of amount of difference.

The basis of building up such a scale is Table XXIII of Thorndike's "Mental and Social Measurements."[1] The theory

[1] The technique involved in this procedure is straightforward. The best references to it are: (a) Thorndike, E. L. "The technique of combining incomplete judgments of the relative positions of *n* facts made by *n* judges," *Journal of Philosophy, Psychology, and Scientific Method*, 13:197-204, April 13, 1916. (b) Knight, F. B. *Qualities connected with success in teaching.* (Teachers College Contributions to Education No. 108 in preparation). (c) Ream, M. J. "A statistical method for incomplete order of merit ratings." *Journal of Applied Psychology*, 5:261-263, Sept., 1921.

is that if 50 percent of a group of judges vote that A is better than B in a certain quality and 50 percent vote that B is better than A, then there is no difference between A and B in respect to the quality. But if A is better than B in the opinion of more than 50 percent of the judges, then A is better than B. The amount of difference depends on the percent of judges voting him to be better. Thus if 75 percent of the judges vote that A is better than B, he is better by 1 unit. If 70 percent of the judges so vote, A is better by 0.78 of a unit. If 98 percent of the judges so vote, A is better by 3.05 units.

Thus by comparing any teacher's ratings with others we can determine her quantitative place in the group. The only arbitrariness here is in giving the worst teacher a value of 1. The one assumption made is that of the competency of the teachers to judge the quality of the work being done. This assumption is justifiable on a common sense basis because after all the good teacher is the one whose associates think he is good, just as the good lawyer is the one whom other lawyers judge to be a good lawyer, or the good doctor is the one to whom other doctors send their patients when they themselves are away. Though teachers rarely, if ever, directly observe each other actually operating in the class, there are many indirect sources of information upon which mutual judgments may be based.

The competency of teachers to rate each other was established statistically in two ways. First, the reliability of their ratings was shown by pooling each of two chance halves of their ratings and correlating these two pooled ratings. This correlation was so high (over +0.9) that ignorance, carelessness, or haste could not have influenced the original ratings to a damaging extent. The second statistical evidence of the worth of teacher's ratings of each other rests upon the fact that these ratings correlated very highly with similar ratings made by supervisors.

Ratings by superior officers.—In addition to the mutual ratings of teachers, ratings from the supervisory staff were obtained in the same way. In Town A, for our example, the supervisors'

rating was a pool of the ratings of the superintendent, four full time principals, and four special-subject supervisors.

Ratings by pupils.—Where departmental work brought three or four teachers before the same class of students, I had the pupils rate the teachers under circumstances calculated to engender a dignified and serious approach to the task. This method was especially useful in the high schools. Here I asked the principals to give me the names of the twenty most dependable pupils in the school. These twenty students rated their teachers. The correlation between the pooled judgments of pupils with that of the supervisors was about +0.7. Some disagreement was present but may as well have been due to lack of insight on the part of supervisors as on the part of students.

The high correlations between the ratings by teachers, supervisors, and (where possible) students made it clear that an order of merit of individuals composing the teaching staffs had been fairly attained. The final quantitative order of merit was a composite of teacher and supervisor judgments. It was easy to be content with one rating, namely, that for "general teaching ability." For the ratings on specific traits agreed so closely with this summarized rating that evidently the general opinion of a rater in respect to the teachers heavily weighted the ratings for distinct traits.

Correlation of Success with Observable and Measurable Facts

With this criterion of teaching success certain objective data concerning the teachers were related. The results will be indicated.

Handwriting.—In another connection samples of the handwriting of every teacher had already been collected. This writing was done under the normal busy conditions of school life. It was in no sense "show" writing but the kind of penmanship customarily used by the teachers. These samples were marked for quality according to the Ayres scale. The correlation between ability to teach and quality of handwriting was found to be zero, *i.e.*, no relationship was discovered.

It would seem as if our prigishness about good penmanship is a bit over-done. A superintendent who discards an applicant's letter because of poor handwriting is using as a criterion of selection a skill which has no evident relationship to success in teaching. This zero correlation of course is not for handwriting in its total range of quality—from stark illegibility to artistic beauty. The range here is quite restricted, for all teachers write fairly well. It is better to say: the differences in quality of penmanship which existed among these teachers bore no relation to differences in teaching skill.

Age.—The age in years and months for all teachers was obtained from the official records of the Massachusetts Retirement Board. The correlation with teaching skill was of negligible size. In one group it was slightly negative, in others positive. We know there is *some* correlation between age in general and teaching ability. A five-year-old child could not teach, and excessive old age would no doubt be negatively correlated. But within those age limits during which men and women ordinarily teach, age does not appear to be correlated with teaching skill. The younger teachers are not the best as a current superstition would lead us to think; nor do years of tenure make material additions to skill.

Experience.—Experience in teaching is not significantly correlated with teaching ability. My data fail, for example, to show that the first five years of teaching are the best or that after that period a teacher begins to slip—notions that are rather widespread among school administrators. The correlation might have been slightly distorted because of the regulations concerning necessary qualifications for new teachers. Towns A and B engage no teacher unless she has had successful experience for two years in other systems. Thus our range is cut off at the lower end. If teachers doing their first or second year of teaching had been at work in these towns some correlation might have appeared. It should also be remembered that teaching is an extremely selective calling. As a teacher gains experience she tends to go into administrative work in proportion as she is possessed of qualities

of leadership and general ability. Superior teachers also tend to go to larger cities where salaries are higher and where the social life is more attractive. These two facts—namely, that no inexperienced teachers were hired and that stronger teachers tend to leave class-room teaching in smaller communities—may help to account for the absence of relationship between amount of experience and amount of teaching skill. At all events, in these towns no correlation exists. Moreover, the probable error due to scanty data is negligible for in this correlation there are over 150 cases.

Professional study during service.—The amount of summer school and Saturday work in educational courses was ascertained for each teacher. A useful relationship between amounts of this kind of study and skill in teaching could not be established. I am confused by this absence of relationship. It would seem that summer school study would be indulged in more frequently by good teachers than by poor teachers. Even if information gained in such study did not function directly in the class room, one might nevertheless hold that a correlation between study and skill in teaching exists. For it would seem reasonable to expect that a teacher who has ambition enough to go to summer school would also tend to be a vigorous and alert teacher in the class room. However no such relationship was found. The zero correlation could not be accounted for by absence of spread in either distribution for we had a good spread in our order of merit in general teaching ability and also in the amount of summer school work.

Mental test.—In Towns A and B all elementary-school teachers were given a mental test one hour in length. This test was the first hour's work of Thorndike's College Entrance Examination, Forms K and J. As would be expected the 120 cases were spread out from very low to very high scores. The correlation with teaching ability in both groups was slightly positive: but it was so low that zero came within the probable error.

It is evident that a mental test is not a good selective device as far as elementary-school teaching is concerned. We should bear in mind, however, that all the teachers involved in this

study were in the upper portion of the total range of intelligence. Obviously an idiot could not teach. These teachers had enough intelligence to go through high school and normal school. What this correlation means then is "given enough intelligence to pass through high school and normal school as a minimum possessed by all," then more intelligence does not seem to function in the practical situation.

Of course one can use any amount of intelligence even genius in elementary-school teaching, but other factors such as interest in children, persistence, *savoir faire*, also operate; and in the general shuffle, differences in intelligence tend to be lost. It is also quite possible that abler teachers leave teaching in elementary schools for supervisory or other work. Too little intelligence would be a handicap. Too much intelligence might also be a handicap. A too active mind might well be so constantly aware of unsolved problems—methods, devices, changes theoretically possible but practically inadvisable—that superior intelligence would get in its own way in the rather humdrum and steady grind of elementary-school teaching.

Normal school standing.—In computing the prognostic value of a student's success in normal school in relation to success in teaching one assumption had to be made which may have been a little risky, namely, that an "A" from one normal school denoted scholarship equivalent to that denoted by an "A" from another normal school. The equivalence of marks, standards, quantity and quality of work of the normal schools cannot be proved. It should be said, however, that this equivalence is generally admitted by school superintendents in New England. Half of the normal schools in my study were supervised from the same office. Frequent conferences, interchange of instructors, and approximately equal support are among the factors tending to make the assumption of equal standards allowable.

Another assumption that was necessary was that marks meant the same thing in different years. Thus an "A" in English in 1906 was assumed to be equivalent to an "A" in English in 1910. As instructors change there is doubtless some variation in the

true meaning of "A" or of any other mark. The value of an "A" in English as compared with an "A" in, say, Art was taken care of by getting all the marks given in any subject and assigning numerical values to the letters in terms of the standard deviation from the central tendency.

To get enough cases of elementary-school teachers with normal-school records complete and from normal schools approximately alike it was necessary to pool the teachers of Town A and Town B. Here we assumed that the median teacher of Town A was equal in teaching ability to the median teacher of Town B. As the spread of merit of these two groups was about alike and as these two towns were similar in important facts such as size, type of community, salary schedules, requirements for beginning teachers, courses of study, and supervisory provisions, this assumption was not at all a violent one. This allowed me to use over 50 cases in the correlation. The correlation between "how well a prospective teacher does in normal school and how well she does in teaching" was practically zero.

This zero correlation means that if a normal-school student does well enough to pass, added excellence in normal-school work does not prophesy unusual excellence on the job. Anyone who is intimately acquainted with elementary school teaching will not be surprised at this correlation. Factors not effected very much by formal study effect very seriously actual teaching success. With a sufficiently minute course of study, with able and persistent supervision, a teacher can soon make up what a "C" instead of an "A" in normal school implies.

Knight-Franzen Trade Test for Elementary School Teachers.—A trade test of elementary school technique and information was constructed. This test is supposed to measure the amount of intimate knowledge concerning elementary school teaching and allied information which is possessed by the individual taking the test. We know that it is not a disguised intelligence test for it does not correlate with scores on intelligence tests. It does measure something related to school teaching for teachers as a group score more highly than do students of equal age and ability

who know nothing about teaching. Thus a group of young elementary school teachers normal-trained and with a year or two experience do much better than a group of advanced law students or students of salesmanship and advertising.

This test was given to the elementary school teachers of Town A and Town B. The correlation between general teaching success and scores on the trade test was +.5. Of all the instruments for estimating teaching success including mental tests, experience, normal school record, summer school study and the trade test, the trade test was the only instrument to turn in a positive correlation of sufficient size to be really useful for purposes of selection. This test is now revised and data on its prognostic value is being gathered in the elementary school group of Des Moines, Iowa.

The findings of this study lead one to wonder how much better than chance the selective skill of the average superintendent is. As far as I know we have on record no correlation between how well a superintendent thought teachers would do and what they actually did. In other words, the correlation between successful candidating and successful performance is unknown. The kind of information usually asked of a candidate does not correlate according to my data with successful performance. In spite of the fact of the complexity of the trait, "general teaching ability," our correlation of +.5 between it and a trade test carries the definite hope for a genuinely scientific procedure of teacher selection in which photographs, conversations, letters of recommendation may weight the decision less heavily and some use of the Trade Test technique be of very definite help.

SOME USES FOR INTELLIGENCE TESTS[1]

SAMUEL S. BROOKS
District Superintendent, Winchester, New Hampshire

Another promising offspring of modern psychological science is the standardized intelligence test. In the hands of practical men intelligence tests are proving themselves to be practical tools for practical purposes. During the war they were used to obtain leaders of men for the army; large industrial concerns are using them to pick young men and women to be trained for executive positions; great universities are using them in lieu of entrance examinations to select students; social welfare organizations are using them to discover feebleminded individuals who menace society as potential or actual criminals; live teachers and educational administrators are using them for various purposes. We have heard much of them during the last three or four years. Are we all getting our share of the help they offer us?

I do not propose to enter into a comprehensive discussion of the nature of intelligence tests, the need for them, their reliability, or the uses to which they may be put. All this has been set forth at length and in a clear and readable manner by Professor Lewis M. Terman to whose book, *The Measurement of Intelligence*, published by Houghton Mifflin Co., I refer all readers desirous of a full discussion of intelligence testing in general and of the Binet-Simon Intelligence Scale in particular. My own purpose is merely to describe why and how we have put intelligence tests to some practical uses in the schools of my district.

I shall, however, refer for a moment to the need for some method of measuring objectively the relative mental abilities of pupils in the schools. This need is not generally felt even among teachers and schoolmen. Some of the teachers, whom I convinced with comparatively little difficulty of the need for standardized tests for measuring the progress of pupils in their studies,

[1] This is the eighth article by Superintendent Brooks on the general topic, "Putting Standardized Tests to Practical Use in Rural Schools."

were inclined to scoff at the idea of intelligence tests. Their attitude may be expressed in the words of one teacher of many years' experience who said with a decided air of assurance, "I guess I can tell the bright children from the dull ones without the help of any intelligence tests." Just before giving the intelligence tests in the school of this particular teacher, I asked her to write down for me the name of the brightest pupil in each grade according to her best judgment.

As might be expected, and as events proved, her estimates were right or nearly right in some cases and entirely wrong in others. As an example of being wrong, she selected as the brightest pupil in grade IV, a twelve-year-old girl of small stature who led her class in achievement. Now it must be perfectly obvious to anyone who considers the matter seriously that it would be a very unusual thing to find a mentally superior child of twelve years, or even a mentally normal one of that age, in the fourth grade. But the teacher had neglected the age factor in making her estimate of this child's mental ability and had rated her as a very bright child simply because she was doing the best work of any pupil in her class. The fact that the child was small for her age and so did not tower above her classmates of nine and ten probably helped out the delusion. If this girl had been placed in a class of normal twelve-year-olds she would have been recognized at once as a dull pupil—a fact which the mental tests at once disclosed. She had a mental age of 10 years, 2 months, and an Intelligence Quotient of 80. In reality the brightest child in the fourth grade proved to be a little girl of 8 years, 6 months whom the teacher estimated as "just average." This girl had a mental age of 10 years and an Intelligence Quotient of 118. Among pupils of her own age she would have been a shining star. The teacher's errors of judgment were due to the fact that because of their ages the first girl was working a little below capacity and the second a little above capacity. Teachers and others are too prone to estimate a child's mental ability by comparing the amount and quality of his work with those of the other children in his grade regardless of how much

he may be advanced or retarded. If the child happens to be in the normal grade for his age, her judgment may be fairly accurate; otherwise it is likely to be inaccurate. Personal judgment in such matters must be replaced as far as possible by scientific measurement.

Furthermore, intelligence tests are needed to help refute a common fallacy which is almost unbelievably widespread in the educational word as well as outside it—the fallacy that under proper conditions and with proper instruction every child barring the obviously feebleminded is about equally capable of making satisfactory progress in any study. This idea is echoed in our Declaration of Independence which offers as a self-evident truth that "All men are created free and equal." It is re-echoed in the rabid mouthings of I. W. W.'s proclaiming the equality of men. It is the precept and guide of the teacher who wastes her time and the time of the brighter children of a class while she holds them back and tries desperately to help one or two mentally deficient classmates keep up with the rest.

Only a short time ago the principal of one of the oldest and most famous academies in New England spent nearly an hour assuring me most vehemently that there was absolutely no reason in the world why, with proper instruction and sufficient interested effort, one student could not do just as well as another in Latin, history, algebra, or chemistry, and this in face of the fact that he admitted he had never been able to achieve such ideally uniform results in any of his classes. I gathered from his talk, however, that the failure was not due in any measure to inadequate instruction, but entirely to widely varying degrees of interest, industry, and application on the part of his pupils. He was cocksure and eloquent. I was so amazed at his attitude and so overwhelmed by a torrent of timeworn, dogmatic and, to him, unassailable arguments upholding his contention, that I could offer but a feeble reply. I am convinced that he went away with the firm belief that I was some new variety of incurable crank.

Now any one with common sense, who will forget proverbs and doubtful platitudes long enough to give his common sense

time to function, can readily see that men are not born equal. Perhaps they should be, but they are not. They are not born equal mentally, physically or financially nor even with that democratic equality of opportunity of which we hear so much. We have all extremes mentally from the driveling idiot to the genius, physically from the bedridden cripple to the physically perfect human being, financially from the pauper to the millionaire. Financial inequality may be more or less overcome, and in many cases so also may physical inequality. But mental inequality, according to the psychologists, seems to be pretty much a fixed condition. They, with their brother scientists in the realm of genetics, seem to have proved to the satisfaction of the majority of their fellows that an individual's mental capacities are determined from the moment of his conception, and that the limitations of his mental development are predetermined by the forces of heredity. This means for us, among other things, that as soon as a school child has reached his limitations, if not before, he will begin to fall behind his classmates who have inherited better mental equipment, and that no amount of extra coaching on our part nor of effort on his part will enable him to keep up for long unless the rest of the class is held down to his pace. If, therefore, efficiency means partly the elimination of wasted effort, should we not, in the name of efficiency, eliminate the waste of time and energy expended in the hopeless task of trying to fit all children to the same mold?

Only for the past few years—in fact, only since the publication of the Binet-Simon Intelligence Scale, in some of its later and better editions—have we possessed a really practical and fairly accurate tool for the measurement of intelligence, one which can be used effectively by interested persons of ordinary intelligence with little experience in psychological testing. This scale—because of the many years of careful investigation and experiment by its originator, the several painstaking revisions, and its careful standardization both as to content and method of procedure—is undoubtedly the most accurate intelligence test available. The fact, however, that each individual must be interviewed separately

makes it unsuitable for general use in testing large numbers of pupils, as, for example, in making an educational survey. For a few examiners to test thousands of children in a large school system, or for one examiner to test several hundred children in a small school system with the Binet-Simon Scale would take more time than is usually available for such purposes. This is especially true if the testing must be done by the superintendent or his assistants along with their numerous other duties.

When we entered the war against Germany our military authorities were faced with the problem of selecting and training thousands of new officers to lead the millions of raw recruits furnished by the draft. Time was at a premium. The psychologists offered their assistance, and after a period of trial it was decided to permit them to select the new officer material by giving intelligence tests to the more promising of the drafted and enlisted men. The above-mentioned limitation to the practical use of the Binet-Simon Scale was quickly realized and led to the hurried devising and standardizing of group intelligence tests by means of which hundreds of individuals could be tested at one time. These tests were to some extent based on the Binet-Simon Scale and to a greater extent on the special mental tests which had hitherto been used in psychological laboratories (directions, analogies, opposites, etc.). The new instruments, however, were adapted in organization and method of procedure to group presentation, definite response, and objective scoring. By the end of the war such tests had reached a high degree of development. The industrial and social world, awakening to the possibilities of mental measurement, soon adopted the army tests for their own purposes, as they could well do since they were dealing mainly with adults. But radical changes were required to fit them for use in testing the mental abilities of elementary-school children. Devising suitable group tests for the younger children who cannot read or write much was the most difficult problem. This problem has been partially solved within the last three years by means of picture completion tests. There are now available a number of excellent group tests adapted to school use.

In this district we have used the Otis Group Intelligence Test for the upper grades, the Dearborn test for the lower grades, and the Haggerty tests for all the grades. Although we have derived several worth-while advantages from their use, the original purpose in giving them was to discover all the mentally incompetent children in the schools of the district. The scheme for measuring the ability of teachers by the progress of their pupils, as described in a previous article (such progress to be measured by standardized achievement tests) demanded some way of finding out which pupils were mentally capable of making somewhere near normal progress and which ones were incapable of doing so. It is manifestly unfair to expect teachers to secure normal progress with feebleminded or very dull pupils. Accordingly, the work of determining the mental ages and intelligence quotients of all the children in the district was undertaken soon after they had been graded in October.

At that time as far as I knew the most suitable standardized group intelligence test on the market was the Otis test. It contains excellent testing material organized and arranged so as to be easily and objectively scored. It is supposed to be used as low as the third grade, but on checking up the results with those from the Binet-Simon Test I found that, although the correlation was fairly high in the seventh and eighth grades, it grew rapidly less in going down the grades until in the third grade it was too small to bespeak much accuracy for the Otis test. This is assuming, of course, that the Binet-Simon Test is the standard in accuracy. Just as a guess, I might venture the opinion that the Otis test makes too great demands on concentration and acquired reading ability to give accurate results below the sixth grade except with the brightest children.

This failure of the Otis test to give accurate results in the lower grades was somewhat discouraging. I had planned to use it with all pupils above the second grade and then gradually, as I could find time, to test out the first two grades with the Binet-Simon Scale. It now appeared that I should have to begin with

the individual tests in the fifth grade and work down. In the interests of uniformity and accuracy I planned to do all the mental testing myself, and even with group tests this would be a considerable task in a district where the schools were so scattered.

Nevertheless I tackled the job and by the end of the Fall term had tested more than seventy pupils with the Binet-Simon Scale. During the first month of the winter term other matters kept me busy. Then came the giving of the midyear achievement tests and the attendant work of tabulating results. I had just got around to taking up the mental testing once more, when the Haggerty tests made their appearance. I ordered some to try out. They were first given, in the four lower grades to the same children to whom the individual tests had already been given. I was delighted to find a fairly close agreement between the Haggerty and Binet-Simon results even in the first and second grades. Then I began all over again, giving the Haggerty tests right through the district in all the grades. The giving of the tests took about a week and by the end of three weeks they were all corrected and the results recorded.

Each pupil's mental age and intelligence quotient were recorded on his or her graph card where they have often proved very enlightening when studied in connection with the pupil's achievement record on the same card. For instance, I have before me as I write, the card of a thirteen-year-old boy with an intelligence quotient of 108. Although his mental age at the time of the test was 13 years, 11 months, he was only in the sixth grade and his achievement record showed that even in that grade he was doing poor work. Now why should a child of his age and intelligence be doing poor work in the sixth grade? Any one or more of various conditions might account for it such as poor teaching, poor general health, adenoids, enlarged tonsils, defective sense organs, unfavorable living conditions at home, constant fatigue from outside work, and so on. But if we are to handle such cases with understanding and sympathy, we must know definitely the cause of the trouble.

In this boy's case it was found on investigation to be a combination of untoward circumstances. To begin with, he was much overworked outside of school hours, often at tasks beyond his strength. Moreover, his parents were constantly quarreling and snarling at each other, making home life a misery to the sensitive boy. A medical examination showed him to be in good general health but revealed the fact that he was a little deaf; a fact never before suspected even by his parents. In connection with this fact it is significant that his teacher was accustomed to speak in rather subdued tones, so that he lost a large part of the oral instruction. To sum up, the boy was hard of hearing, sensitive, tired, and discouraged.

Having discovered these conditions, we moved to remedy them as far as possible. Seating the boy where he could watch the teacher's lips at all times when she was talking to the class enabled him to get much instruction which under former conditions would have been lost to him. A tactful show of sympathy and understanding on the teacher's part and words of encouragement instead of constant nagging for failure to do the class work brought a new light to his eyes and the sullen look of a misunderstood boy left his face. His whole attitude toward the school and its work changed. A talk with his father, who did not mean to be either unreasonable or unkind, helped to lighten his burden of work at home. A talk with both parents concerning the effects of their constant bickering on their boy's life, served to make home life more pleasant. They were really a devoted couple and their quarreling seemed to be more from habit and more because they enjoyed it than because of real ill-feeling. This pupil is undoubtedly a much happier boy, interested in his school work, and gradually catching up with the other children of his age and ability in the school. He is no longer considered dull.

Such investigations and readjustments are decidedly worthwhile. But before they can take place our attention must be attracted to the need for investigation. And here the worth of standardized tests, intelligently used, is again demonstrated. In this boy's case it is interesting to conjecture how long it might

have been before the need for investigation would have become apparent, if the results of his mental and achievement tests had not been recorded on the same card and carefully studied together by someone interested in interpreting them for the best good of all concerned.

This was a particularly interesting case, and for this reason it was chosen to illustrate my point. But it is not the only case where comparison of mental test records with achievement test records has led to investigations resulting in permanent good. And there is need of investigation whenever a child grades high in mental ability and low in actual achievement of school work. For it is very unusual to find a pupil, physically and mentally normal, in good health, and with good home influences, who is doing unsatisfactory work in school. If such is apparently the case, there is generally something wrong somewhere; and it is usually possible to make a beneficial readjustment.

Another card represents a type of pupil constituting one of the serious school and social problems. It is the record of a boy of 15 years, 6 months. His mental age is 9 years, 2 months and his Intelligence Quotient, 59. This boy's graph shows that he cannot do satisfactory work in the third grade, although he has been in that grade for four years. Investigation revealed his immediate ancestry to be of unusually low mental and moral caliber. This boy is a menace to the school and the school is a menace to him. He is a menace to the school because, with all the dawning strength, instincts, passions, and emotions of the primitive male controlled only by the undeveloped mind of a nine-year-old, he is no fit associate for normal boys and girls. The school is a menace to him because, instead of furnishing him with interesting and valuable employment suited to his abilities, it is wasting time that he could more profitably and instructively employ elsewhere and because it is forcing upon him habits of idleness and failure. Unless special classes are available, the public school is no place for children with intelligence quotients much below 70 especially when they have become two or more years retarded in their school work. No further evidence is needed that they have reached the

limit of their mental development along the lines of the ordinary program of studies. They should be transferred to special institutions whenever possible where they can have special training suited to their needs and capacities, and measures should be taken to prevent them from reproducing their kind. At any rate, the public schools should be rid of them. With the proofs furnished by the results of intelligence tests, backed by the child's record of achievement in his school work, it ought to be possible to get school boards to act in excusing such children from school even if they cannot be otherwise properly taken care of.

Still more of a problem, from all points of view, are the children with intelligence quotients ranging from 65 to 80. More often than otherwise they appear superficially to be entirely normal or even bright. They may do excellent work in the first four grades where habit formation is the chief end to be attained and where drill is the chief feature of instruction. Sometimes they continue to do fairly well even in the upper grades in schools where much rote memorizing prevails and where memorizing ability is mistakenly considered an index of general intelligence and learning power. But in properly conducted schools such children begin to fall behind their classmates in the fifth and sixth grades and soon become hopelessly retarded. They have reached the limits of their abilities in learning from books or from ordinary schoolroom instruction. They are very much lacking in the powers of initiation, discrimination, and reasoning demanded by the higher types of learning. They are incapable of higher thought processes. Hence they fail in grammar, problem-solving in mathematics, and in the content subjects if the teaching of the latter demands, as it should, more than mere memorization of facts.

Teachers are often unjustly blamed because such children fail in their work. Parents wonder why their children cannot learn under the new teacher as well as they did under the old when it is in no way the fault of the teacher. The children have simply reached their limit of mental development. If a child reaches the limit of his mental development at a mental age of eleven years

he will never be much older than eleven years mentally, though he lives to be a hundred. Children of this sort are too often permitted to become the pacemakers in their classes to the untold harm of the brighter pupils. But in spite of all attempts to keep them along with the other and brighter children of the same age, they finally get completely beyond their depth and fail day after day in their school tasks until they begin to believe they are absolute failures and that success in anything is impossible for them. They grow discouraged, give up trying, and devote themselves to mischief or wait passively for the legal age limit to be reached so that they can leave school and go to work at something where they can at least earn a little money in return for their time. Meanwhile, failure and idleness have all too often become habits that follow them beyond the schoolroom and lessen their social efficiency. In such cases the school has defeated its own ends.

This type of child, the high-grade moron, is such a problem partly because he is not generally recognized as being mentally deficient. The shortcomings of the feebleminded are generally evident, and no one expects much from them. The high-grade moron, however, is usually normal in appearance and in ordinary intercourse with other people appears to be normal mentally. It is only when situations arise which demand the functioning of the higher forms of intelligence that he reveals his deficiencies. Even his teacher often fails to understand why he continually fails in his school work. He is characterized as "obstinate" and "lazy" when in reality he is mentally deficient and incapable of doing the work demanded of him. Carefully conducted mental tests will reveal such cases and should lead to more sympathetic and intelligent treatment of them.

When such children become retarded as much as two years, it is little less than criminal to keep them in rural or small-town schools where there are no special classes for their benefit and where there is not sufficient differentiation of courses to permit of their being given amounts and kinds of work suited to their abilities. If possible they should be sent to manual trades schools where they can be taught a trade and at the same time be given

as much cultural training as they are capable of acquiring. Otherwise some arrangements should be made whereby they can leave school and go to work under the supervision of their parents or of other responsible persons who will instruct them in the rudiments of some useful line of work. They might still be under the supervision of the school authorities to the extent of being obliged to spend as many hours per week in useful labor under real instruction as they would normally have to spend in school; such supervision of the school authorities to continue until the legal age limit for compulsory attendance is reached. And why could not school credits be allowed for such work?

To return, however, to my main purpose in giving the intelligence tests. When I first discussed with the teachers of the district the feasibility of some scheme of rating teachers based principally on the progress made by their pupils, it was objected that the varying mental abilities of the children would make such rating unfair unless the records of the slower pupils were ignored. But who was to be the judge as to which pupils were incapable of making normal progress? It was this situation which led us to make use of the intelligence tests. We agreed to discard, in calculating the teachers' rating the records of all pupils with intelligence quotients below 80. Later a still better scheme was worked out based on the average intelligence quotients of the different schools in such a way as to take full cognizance of the varying mental abilities of pupils. Thus, a teacher with a school composed in general of dull children would not suffer in comparison with a teacher of equal ability with a school largely consisting of bright children. That is, two teachers of about equal ability would get approximately equal ratings regardless of the comparative average mentalities of their respective schools. In this scheme the intelligence test was the impartial judge whose findings were accepted as satisfactory by both teachers and superintendent.

Another valuable service of the intelligence tests was in locating the children of very superior abilities. In one school of 25 pupils I discovered five with intelligence quotients around 140

and none with intelligence quotients of less than 80. This school was locally noted as being uniformly fortunate in securing good teachers under whom the pupils advanced very well indeed. Of course it would be a mighty poor teacher that couldn't get passable results with such a school. In another school I found, working with other children of about the same chronological age, a little girl of 8 years, 6 months with a mental age of twelve years and an intelligence quotient of 142. In the ordinary run of events she would probably have secured no further recognition of her superior abilities than regular promotion and graduation in due time. She is now 9 years, 5 months old and leading her class in the fifth grade. She could probably do satisfactory work in the sixth grade.

This type of pupil is not generally recognized as a serious problem. And in truth, the situation is more serious to the pupil and the public than to the teacher, however unconscious of the fact they may be. As at present organized, the average school is probably doing such children as much harm as it is good, in that it does not furnish them with opportunity and incentive to develop their capacities to the limit. It is from among these children that the leaders of the future are to come, and the public which foots the bills is being cheated when such children are not given opportunity to develop as they should. The present generation is retarded in progress by the provincial narrowness of natively intelligent but mentally blind leaders of the blind whose possibilities for development were early cramped to deformity in the narrow confines of the traditional elementary- and secondary-school programs by being forced to travel in a groove in competition with their mentally inferior classmates. Our school system can never attain nearly its maximum efficiency until programs and courses of study are so differentiated as to fit the kind and quantity of work to the mental capacities of individual pupils. The standardized intelligence test will prove a useful tool in shaping plans toward that end.

The problems noted above are present to a greater or less extent in practically all school systems; but it is in the cities

that, with the help of standardized tests, they may be most easily solved. The city superintendent has, or can arrange to have, special classes or even special schools to take care of the feeble-minded children so as to separate them from the other pupils. Those troubled with sensory defects or other serious physical deficiencies can be placed in special classes or institutions provided for their benefit. Then with two, or preferably three, divisions in each grade, the children can be assigned on the basis of mental ages and intelligence quotients to groups from which fairly uniform work can be expected.

Let us suppose for instance, that intelligence tests have been given to the fifth grade in a large city school and that the mental ages and intelligence quotients of all the pupils have been recorded on cards together with their respective names. First, the average mental age for the whole grade should be computed and all cards sorted out which are marked with mental ages more than a year above or below the average for the grade. The pupils represented by these cards should be given special attention. Their achievement records should be studied and, where possible, they should be retested with the Binet-Simon Test if that test was not used in the first instance. It will probably be found that most of these pupils do not belong in the fifth grade at all but in the fourth or sixth grade. Next, the remaining cards may be sorted into four piles. Let the first contain the intelligence quotients below 70; the second, those from 70 to 90; the third, those from 90 to 115; and the fourth, those above 115. These groupings are of course rather arbitrary and might be varied somewhat. All pupils with intelligence quotients much below 70 would probably be best placed in special classes or schools. The other three groups might be termed the A, B, and C divisions of grade V; and the children in each group, being of approximately equal mental age and intelligence, should be able to do about the same kind and amount of school work. Since there are no sharp lines of demarcation between groups, some further adjustments would probably be made as experience might dictate.

The courses of study could then be adapted to the various groups. The group of dull pupils might be assigned a minimum

of book work and a maximum of manual training. The pupils in the normal group would perhaps divide their time somewhat evenly between these two lines of work. The superior group could probably accomplish as much of the manual work as the normal group and much more of the mental work.

In this way intelligence tests may assist in bringing about that much needed differentiation in courses of study which will permit each pupil to have work more suited to his particular abilities. Although it is advisable, where possible, to have a double grouping based on the results of intelligence tests—namely, one grouping according to mental ages for classification by grades and the other grouping based on intelligence quotients for subdivisions of grades—such a double grouping is not usually possible in the smaller schools where all the pupils in a grade must work together. In these cases the grouping must depend more on the mental ages, as will be explained further on.

It was while engaged in recording the mental ages and intelligence quotients of the pupils on their respective graph cards that the possible use of intelligence tests for grading purposes occurred to me. I gradually became conscious of the fact that, although there was wide variation in the chronological ages of the children in any one grade (as they had already been graded by the standardized achievement tests), the mental ages in a grade did not usually appear to vary by more than a few months, or a year from the average mental age for the grade. If, therefore, as may apparently be expected, there is close correlation between the results of intelligence tests and the combined results of achievement tests, that is, between mental ability and accomplishment of school tasks, why cannot intelligence tests be used instead of achievement tests for grading purposes, at a considerable saving of time and energy?

When, therefore, I had finished recording the results of the intelligence tests on the graph card, I proceeded to develop this idea. The chronological and mental ages of all the sixth-grade pupils in the Tamworth schools were tabulated as shown in Table I.

The variation in chronological ages within this sixth grade is from 9 years 5 months for the youngest pupil to 15 years 3 months for the oldest—a total range of 5 years 10 months, or of nearly six years. On the other hand, the variation in mental ages of the same children is from 11 years 5 months to 13 years 5 months—a total range of only two years. If the first two pupils and the last pupil are omitted the range is only one and a half years. These facts would indicate that the grading with the achievement tests had served to bring together children of much the same mentality and that intelligence tests would have served equally well if not better for grading purposes. Similar tabulations were made for the second- and third-grade pupils. The results were not quite so convincing but were still encouraging. The range in mental ages for the third grade was 1 year 9 months and for the second grade 2 years 3 months. There was 22 percent of overlapping of mental ages between these two grades. The increasing range of mental ages within a grade, as

TABLE I.—COMPARISON OF MENTAL AND CHRONOLOGICAL AGES OF SIXTH-GRADE PUPILS. TAMWORTH SCHOOLS

Chronological Ages		Mental Ages	
9–5	12–3	11–5	12–4
9–10	12–3	11–6	12–5
10–8	12–5	11–8	12–6
10–10	12–5	11–8	12–6
10–10	12–6	11–9	12–6
11–0	12–7	11–10	12–6
11–5	12–9	11–10	12–6
11–6	12–10	11–10	12–9
11–7	12–10	12–0	12–10
11–7	13–2	12–0	12–10
11–8	13–7	12–0	12–10
11–8	13–6	12–0	12–10
11–11	13–9	12–1	12–11
12–0	14–0	12–1	13–0
12–0	14–4	12–2	13–0
12–1	14–8	12–2	13–2
12–1	14–8	12–3	13–2
12–2	15–2	12–3	13–2
12–3	15–3	12–3	13–5
	15–3	12–4	

we go down the grades, would seem to support the reasonable and suspected fact that all tests are less accurate and dependable the younger the children tested. The overlapping encountered in the two lower grades led me to make still another tabulation—namely, of the seventh grade—so that I could compare it for overlapping of mental ages with the sixth grade. There was an overlap of about 14 percent between these two grades. This would indicate that the amount of overlapping also increases from higher to lower grades and probably for the same reason mentioned above for the increase in range of mental ages.

Then I began to wonder how much differently the pupils would have been grouped by grades if the grading had been done with intelligence tests instead of achievement tests. So I constructed a distribution as shown in Table II. Having no established data as to what the limits of the range should be for the different grades, considerable experimenting was necessary with different arrangements of intervals. A two-year range was finally adopted since I had already found, as described above, that the range of mental ages for each grade was about two years. Since children ordinarily enter school between ages six and eight, and since the chronological age of a normal child corresponds with his mental age, the first grade was assigned the range of mental ages from 6–0 to 7–11. (The expression 7–11 means 7 years, 11 months and so for like expressions). The interval used in Table II is one year, but two intervals are allowed for each grade in order to show a more exact distribution. This arrangement also allows for an overlapping between grades of one year of mental age. Thus the normal range for the first grade pupils is from 6–0 to 7–11, that for the second grade from 7–0 to 8–11, that for the third grade from 8–0 to 9–11, and so on. Moreover, the children in any annual age group may be normally in either of two grades. For instance, the nine-year-olds are normally placed in either the third or fourth grade. The first grade might be regarded as having a range of three years because all children with mental ages below 6–0 will naturally be included in that grade. A somewhat similar statement may be

made for the eighth grade, since it might be expected to include any mental age above 13–0. This bunching of mental ages is, of course, due to the fact that these grades are the lowest and highest respectively in the school.

TABLE II.—MENTAL AGES OF PUPILS IN EACH GRADE

Mental Ages	Grades								Total
	I	II	III	IV	V	VI	VII	VIII	
5 to 5–11	8								8
6 to 6–11	39	4							43
7 to 7–11	45	43	4	1					93
8 to 8–11	2	46	28	5					81
9 to 9–11		5	41	26	4	1			77
10 to 10–11			3	19	27	6	1		56
11 to 11–11			2	4	36	23	5		70
12 to 12–11				1	4	47	18	3	73
13 to 13–11					2	3	28	3	36
14 and over						6	15	34	55
Total	94	98	78	56	73	86	67	40	592

Table II shows the distribution of pupils as they are actually graded in this district on the basis of achievement in standardized tests. Five hundred and ninety-two children are represented. All figures between the heavy zigzag lines represent children normally placed according to their mental ages. The figures outside these lines represent pupils working, for some reason, in higher or lower grades than the tests show them to be fitted for. Take the third-grade column for instance. In this grade there are 28 children between 8–0 and 8–11 and 41 between 9–0 and 9–11 that rightly belong there. There are four children in the grade that, according to their mental ages, ought to be in the second grade, and on the same basis still other children belong in the fourth and fifth grades.

Now, if all the schools were graded strictly on the basis of the mental ages of pupils, *all* the figures would fall between the zigzag lines. Hence 89 out of 592 or 15 percent of the pupils are working above or below their mental capacities. Of these, 7.1 percent are working above or trying to, and 7.9 percent are

working below. This shows on the whole a pretty close agreement between the results of intelligence tests and achievement tests and indicates that either gives much the same basis for grading.

I have studied the records of the children whose mental ages show them to be misplaced and find that most of those working above normal grade are pupils that had been pushed too far ahead by their teachers before the preliminary grading took place and whom, since they were working hard to hold their places, we did not demote when regrading. About half of the pupils shown as working below their apparent mental capacities are accounted for by retardation due to sickness or absence, poor general health, sensory defects, etc. Most of the others remain a mystery as yet. Gradually we hope to work out most of the misplaced ones into their proper mental-age groups so that the grouping of pupils throughout the district will be based pretty closely on mental age.

City schools can be graded with intelligence tests into eight mental-age groups as shown in Table II and then, since there will be a wide variation of intelligence quotients within each mental-age group, these groups can be subdivided into grade divisions or sections as explained above. But the smaller rural schools without grade subdivisions must necessarily be satisfied with a less perfect distribution. Let us suppose that we have given the intelligence tests in a rural school and that the mental ages and intelligence quotients of the pupils are found to be as shown in Table III. Now, let us select all the mental ages within the first-grade range (all below 8–0); and let us list them separately in Table IV with their intelligence quotients.

In discussing Table II it was noted that part of the children in the 7–0 to 7–11 age group would normally be in the second grade. How shall we decide which ones belong in the second grade? To begin with, no child just beginning school will belong there. Accordingly, in Table IV let us mark with a star the number of each pupil who attended school the previous year.

These pupils may enter the second grade if their mentality is normal or better. Let us therefore mark with an additional star those with intelligence quotients of 90 or above. Those thus marked with a double star may be assigned to the second grade.

TABLE III.—RESULTS OF MENTAL TESTS IN A SELECTED SCHOOL

Pupil Number	Mental Age	Intelligence Quotient	Pupil Number	Mental Age	Intelligence Quotient
1	5–0	76	23	7–0	80
2	7–1	76	24	13–0	106
3	7–11	103	25	14–7	110
4	7–9	79	26	9–4	73
5	6–11	100	27	7–9	83
6	8–4	91	28	9–8	80
7	15–7	128	29	7–8	68
8	10–8	100	30	14–0	110
9	15–0	123	31	9–10	78
10	11–10	84	32	9–0	93
11	13–7	108	33	11–0	108
12	10–9	132	34	12–4	80
13	9–2	96	35	9–0	95
14	11–1	100	36	8–0	92
15	13–9	138	37	10–6	85
16	9–7	101	38	10–1	73
17	8–5	71	39	8–6	100
18	10–0	90	40	9–1	85
19	7–9	95	41	12–0	83
20	10–6	70	42	13–9	102
21	13–5	108	43	8–7	73
22	8–6	82			

TABLE IV.—PUPILS WITH MENTAL AGES BELOW 8 YEARS

Pupil Number	Mental Age	Intelligence Quotient
1	5–0	76
2	7–1	76
3**	7–11	103
4*	7–9	79
5	6–11	100
19**	7–9	95
23	7–0	80
27*	7–9	83
29*	7–8	68

TABLE V.—PUPILS WITH MENTAL AGES BETWEEN 8 AND 9

Pupil Number	Mental Age	Intelligence Quotient
6*	8–4	91
17	8–5	71
22	8–6	82
36*	8–0	92
39*	8–6	100
43	8–7	73

TABLE VI.—PUPILS WITH MENTAL AGES BETWEEN 9 AND 10

Pupil Number	Mental Age	Intelligence Quotient
13*	9–2	96
16*	9–7	101
26	9–4	73
28	9–8	80
31	9–10	78
32*	9–0	93
35*	9–0	95
40	9–1	85

TABLE VII.—PUPILS WITH MENTAL AGES BETWEEN 10 AND 11

Pupil Number	Mental Age	Intelligence Quotient
8*	10–8	100
12*	10–9	132
18*	10–0	90
20	10–6	70
37	10–6	85
38	10–1	73

All the others will therefore go to make up the first grade.

Next let us consider the mental age group from 8–0 to 8–11 listed in Table V. Part of these belong in the third grade and part in the second. We will choose those with intelligence quotients of 90 or better for the third grade and mark their numbers with a star. The second grade will then consist of pupils 17, 22, and 43 from Table V and pupils 3 and 19 from Table IV.

Table VI lists the pupils of the mental age group from 9–0 to 9–11. Those with starred numbers (intelligence quotients of 90 or above) will be placed in the fourth grade while the rest, together with those having starred numbers in Table V will make up the third grade.

Table VII lists the pupils of the age group from 10–0 to 10–11. From this table pupils 8, 12, and 18 will be placed in grade V while pupils 20, 37, and 38 will go with pupils 13, 16, 32, and 35 from Table VI to make up the fourth grade. There may be some question as to the advisability of placing pupil No. 18 in the fifth grade since his intelligence quotient is barely 90. On the other hand, if pupil No. 12 has had good instruction and is ambitious and industrious he may be able to work in grade VI since his intelligence quotient of 132 shows him to be a pupil of very superior ability. The remaining children listed in Table III may be assigned to their grades in like manner, whereupon the preliminary grading of the school is completed. Some readjusting will probably be found necessary for various reasons which the results of intelligence tests do not allow for. But when once adjusted there should be possible a uniformity of achievement within grades that is absolutely impossible in the school as ordinarily graded where pupils differing in mental age by as much as five or six years are often found working together, or trying to work together, in the same grade.

As I have said before, when I first attempted to grade pupils there were available no standardized group intelligence tests that would cover all the grades. I therefore used achievement tests for grading purposes, as explained in my third article. Although that grading has proved satisfactory, and although Table II plainly shows it agrees closely with the grading that would have been made if intelligence tests had been used, nevertheless, if I had any more preliminary grading to do, I should do it with the group intelligence tests supplemented by the Binet-Simon individual test wherever there was any doubt of the accuracy of the results of the group tests.

THE VALIDITY OF THE WHIPPLE GROUP TEST IN THE FOURTH AND FIFTH GRADES

HELEN DAVIS
Director of Measurements and Special Education, Jackson, Michigan

Among the special classes in operation in the schools of Jackson are what are known as "speed" classes, wherein pupils accomplish two semesters' work in one semester. Candidates for these classes are selected in part by means of group intelligence tests. In the spring of 1921 it was felt desirable to supplement the National Intelligence Test, Scale A, with some other group test in order to make a more accurate selection of pupils for the speed classes of the following autumn. The Whipple Group Tests for Grammar Grades were originally designed for the specific purpose of selecting gifted pupils—a purpose which was also ours. The Group Tests for Grammar Grades, however, had primary reference to grades above the fifth. Their applicability in the fourth and fifth grades was experimentally investigated with results which will be reported in this paper.

PRELIMINARY TRIAL

In order to see whether the tests would operate smoothly and whether the scores obtained would accord approximately with estimated intelligence, a preliminary trial was made with 55 pupils in two rooms that had previously been tested with the National Intelligence Test and in which the pupils were known to be on the whole of about average ability. The pupils belonged to the IV-B, IV-A, and V-B grades.[1]

Changes in instructions.—In this preliminary trial it was found necessary to alter the instructions somewhat, especially in the lower half of the fourth grade. In substance, these alterations were as follows:

Tests 1 and 2. No changes.

Test 3. Because some pupils were unfamiliar with the terms 'plus' and 'minus' the directions were changed to: "If the sentence is true, put a little cross in front of it (illustrating on the blackboard); if the

[1] At Jackson "B" indicates the lower half and "A" the upper half of each yearly grade.

sentence is not true, put a dash, or a little straight line in front of it (also illustrated)." The test was not begun until it was certain that all the pupils had marked correctly the four sample sentences. Finally, there was added to the directions a warning not to ask questions and to skip a sentence that could not be understood—this because some of the pupils could not recognize in print words, like "kindergarten," that they knew orally, and because others interrupted the testing to ask such questions as: "How far is it from New York to Chicago?" (No. 10) or "How many pounds are there in a ton?" (No. 17).

Test 4. A warning was added not to ask questions.

Test 5. The instructions were amplified, especially in the fourth grade, substantially as follows: "Let us suppose that the square at A is a sheet of paper. Then it is folded like this (illustrated with an actual sheet of paper), so that it looks like B. After that, a hole is punched clear through it, as at C. When the paper is unfolded, it will have a crease in it where the dotted line is at D. How many holes will there be in it? Draw little circles to show where they will be (illustrated by drawing such circles on the blackboard in a figure like D)." After concluding the directions as given in the test blank, there was added the final caution: "Remember to read the directions and do exactly what it says to do."

Test 6. The instructions were considerably amplified, substantially as follows: "Two sentences may use different words, but *mean* the same thing. Look at the first sentence under English Proverbs (pointing to heading on the blank). It says: 'Don't shout until you are out of the woods.' Now look at the column of Arabian Proverbs. The second one says: 'Don't throw away your stick until you get to the top of the hill.' That *means* the same thing as the English proverb we just read. It has a number 2 in front of it, so there is a number 2 in front of the English proverb to show that it has the same meaning. Now look at the second English sentence, or proverb: 'A poor cask will hold good wine.' Think what it means; then go to the Arabian Proverbs and hunt for one that means the same thing. When you have found it, see what number is in front of it, and write that number in front of the English proverb. Go ahead and do it." When certain that all the pupils had done this correctly, we continued: "Now find what number should be written in front of the last proverb, and write it there." When all the pupils had done this, the instructions for work on the test itself were added thus: "On the next page are more proverbs, one set of English and one set of African proverbs. You are to 'match up' the proverbs that mean the same thing. Look at the first proverb in the first column. Think what it means, and then hunt in the second column for one that means the same thing. When you have found it, write its number in front of the one in the first column, just as you did on this page. Be sure to find the right number for each proverb."

Individual help may still be needed for some low-fourth grade pupils, after the others have begun the work.

Results.—The tests, when administered with these alterations in the instructions, yielded satisfactory distributions and accorded well with the teachers' estimates of general intelligence. Without pausing to show the entire distributions, the general drift of the scores may be shown as follows:

Grade	Minimum	Median	Maximum
IV-B	20	48	72
IV-A	24	52	96
V-B	40	56	104

Testing the Selected Pupils

Each teacher of grades IV and V throughout the city was directed to select the best five pupils in her grade, and to send them, on a given day, to a central building for examination.

The results of the examination of the 142 pupils thus selected by the teachers are indicated in Table I.

TABLE I.—DISTRIBUTION OF SCORES IN THE WHIPPLE GROUP TEST FOR GRAMMAR GRADES

Score	Grade IV-B	Grade IV-A	Grade V-B	Grade V-A
40–44	3	0	0	0
45–49	3	0	1	0
50–54	5	2	1	0
55–59	2	7	0	0
60–64	4	4	0	0
65–69	1 (Med.)	3	1	0
70–74	2	3	3	2
75–79	2	6 (Med.)	5	2
80–84	3	5	2	5
85–89	3	5	2 (Med.)	4
90–94	0	4	6	7 (Med.)
95–99	1	5	1	3
100–104	0	0	2	4
105–109	0	0	3	3
110–114	0	0	0	4
115–119	0	0	0	5
120–124	0	0	1	0
125–129	0	0	2	0
Total	29	44	30	39

Selecting for the "Speed" Classes on the Basis of the Test Scores

From the 142 pupils here tested and from the unselected group of 55 tested in the preliminary trial of the Whipple tests, about 60 were to be chosen for the speed rooms.

For admission to the speed rooms, three factors were considered: (1) chronological age; (2) the judgment of the teachers, already at hand in the list of names of the best five pupils in each room, arranged in order of rank in their class; and (3) the reaching or surpassing of a certain critical score in the intelligence test.

For the determination of the critical scores in the Whipple test, data from other cities showed that for unselected sixth-grade pupils the 50-percentile is 81, and that for unselected seventh-grade pupils the 40-percentile is 94 and the 50-percentile is 101. On this basis, the following standards were set as criteria (combining factors 1 and 3 above) for admission to the speed rooms from the fourth and fifth grades:

Grade IV: a score not less than 80 and a chronological age not over 11 years.

Grade V: a score not less than 93 and a chronological age not over 12 years.

The effect of these standards is to select from the fourth grade, pupils at least equal to the average sixth-grade pupil, and to select from the fifth grade, pupils who were practically (fortieth percentile) equal to the average seventh-grade pupil.

The pupils who met these standards began work in the speed classes in September, 1921, and at the present writing (November) their work has proved unusually satisfactory. Watch is being kept of their progress and an attempt will be made to study the cases in which progress in the speed room does not seem to justify the selection.

Correlations Between Tests and Between Tests and Scholarship

Scores on the National Intelligence Tests, Scale A, Form I, were available for 174 of the pupils who took the Whipple Group Tests for Grammar Grades. The correlation between these two test scores was 0.74.

Correlations were also computed between each of the general intelligence tests and the scholarship records of the pupils in the unselected group (the pupils used in the preliminary trial). To make these correlations possible, it was necessary to translate the school marks from letter ratings to numerical equivalents and to express the intelligence scores of the two school grades on a common basis. For the first purpose we took the pupils' standings for the preceding semester in reading, spelling, arithmetic, language, geography, and history and gave a credit of four points for each A, three points for each B, two points for each C, and one point for each D. The total of these points was taken as the scholarship standing. For the second purpose, the intelligence test scores for each group were distributed in such a way that the percentile equivalent for any score could be ascertained. The scores obtained by IV-B pupils were then raised to an equivalent IV-A score, and those obtained by V-B pupils were reduced to an equivalent IV-A score. Thus, for example, in the Whipple Group Test a IV-B pupil who scored 54 was found to rank at the 70th percentile for that grade; he was then, for the purposes of the correlation, considered as having scored 61, which was the score at the 70th percentile for the IV-A grade.

The results show a correlation of 0.64 between scholarship and the National Intelligence Test scores (49 cases) and a correlation of 0.68 between scholarship and the Whipple Group Tests scores (54 cases).

Conclusion

The Whipple Group Test for Grammar Grades, originally designed for selecting gifted pupils in the upper grammar grades, can be successfully employed for the same purpose in the fourth

and fifth grades, provided some amplifications are made in the instructions for three of the tests. The results obtained from these tests, when thus modified, show a very high correspondence with teachers' estimates of intelligence, with scores for the National Intelligence Test, and with scholarship records, while the pupils who are selected by them appear to form a very satisfactory group in such speed classes as we are operating at Jackson.

Editorials

REGULATING THE PROJECT

The history of American education is a chronicle of fads which have been hailed by the thoughtless as panaceas for all educational ills and have, in a short time, disappeared. But no fad of major rank has failed to have some contribution behind it. The object lesson which swept the country in the Golden Age of Oswego has left the language lesson as its contribution. The Herbartians, whose Five Formal Steps were hailed as the form of all instruction, trained the American teacher in the art of lesson preparation. The rational method, with its disciplinary extravaganza of moral suasion and its instructional principle of explanation before drill, was the matrix out of which has developed the so-called rational methods now in use. Interest is convalescing from its emotional measles and is getting ready to take its place through definition and limitation among the staple concepts of educational theory.

The latest idea to seize the center of the stage is the project method. At present it is the shibboleth of the progressive. The term is being warped from its original meaning as it passes from school to school across the land and there is no common definition by which to describe its varying forms as it appears in practice. Whole subjects are said to be taught upon the project basis and some schools have developed their curriculum entirely around what they call projects.

Though this can be expected to go on for some time yet, we know that some very important contributions will be left after "the hurly burly's done." Every extreme has its day and after that comes the roughhouse of the initiation ceremonies, when it is being shown its place before entering upon its long career of stable usefulness in chastened spirit born of criticism caused by failure to make good on all its claims.

It is highly desirable that this period of initiation be shortened, and this can be done by judicious planning. The many friends and well wishers of the method should recognize its limitations and so speak, write, and practice, that the stage of stability may be reached as soon as possible. Specifically, they should decide upon the fundamentals which the schools are supposed to teach to children; for there is some body of fundamental experience which can be accepted upon either traditional or objective grounds. Projects may come and projects may go, but this body of experience must be taught. It is the part of valor and of wisdom for the project curriculum maker to see that he does not merely teach projects but that he assures himself before meshing them into a going production unit that they are varied enough in form and content to give the fundamental experiences, which must be taught, the right proportion of emphasis and instruction. He should even be willing to engage a "mopping up" battalion or two to consolidate the instructional ground gained by his shock battalions. Without the necessity for apology, the project can be accompanied by a systematic study of subjects, by drills, and by exercises. The project is a wonderful agency for gaining ground. But it is too emotional and too tense to have the patience to consolidate experience through drill and skill and refinement of method.

Teachers can render significant service to the project method by taking a critical and sympathetic attitude toward it, with the full expectation that it will have a permanent contribution, but with the determination to see that its period of exuberant extravagance is shortened by an enlightened recognition of its contribution and its limitations.

W. W. C.

THE BUILDING PRINCIPAL

Superintendent Nelson's story at the N. E. A. meeting in 1909 has often been repeated. An enthusiastic young teacher returning home for the holidays describes her work to an interested audience. Pupils, patrons, and fellow teachers—even the janitor—pass in review. Finally some one asks the young lady

about her principal. "Oh, the principal," says she, "why, do you know, I hadn't thought of him because in my experience I have come to think of the principal as the man who hands out the chalk." Here and there in the scanty literature bearing upon the status, duties, and responsibilities of the principal bobs up Superintendent Nelson's chalk story.

Yet logically and historically the building principal occupies a commanding position—logically because the building is the natural supervisory unit, and historically because he is "the single time-honored official found in the modern school system." He appeared upon the scene long ago, and compared to him the superintendent is a belated arrival. His power is not as extensive as that of the superintendent, but it is more direct; it is not as intensive as that of the teacher, but it is more influential.

If, however, the principal's position is commanding, it is also so beset with hindrances that its importance in recent years is often potential rather than actual. It is "perhaps the weakest place in our city organization and administration." Yet most of its unrealized possibilities are the result of the system—of that highly organized hierarchy which characterizes most large city systems. It is not necessary, far less is it desirable, that the advent of the superintendent with his assistant and district superintendents and his staff of functional experts should rob the principal of his original jurisdiction or impose petty requirements upon him. Nevertheless such has been the case.

At New York, where the hierarchy to which we have referred is especially in evidence, the by-laws of the Board of Education name—or did name at the time of the survey in 1912—twenty-four distinct duties of the principal. Most of these duties have to do with reporting something. They include the preparation of requisitions, the keeping of inventories, the care of school property, the supervision of janitors, the conduct of fire drills, the preparation of payrolls, and attention to the absence and lateness of teachers. A list of these tasks gives one the impression that the fundamental duty of the principal is to make as many different kinds of reports as the central office can devise.

Nothing, however, inheres in his position which precludes the exercise of a predominant influence in the direction of better training for children. Observe, if you will, an organization chart such as that which Cubberley presents for a city employing 350 to 400 teachers. Note how the lines converge toward and diverge from the school principal. The city superintendent and his assistant superintendents operate through him. The superintendent of properties, the business manager, the attendance department, the school efficiency man, the health director, and the city librarian bring their influence to bear upon teachers and pupils largely through him. His authority is exerted not only upon the teachers, both regular and special, but also upon janitors and school nurses. It is he who, more than any other official, comes into direct contact with parents. Visitors, whether they belong to the system or not—even the superintendent himself—defer to the authority of the principal while in his building. He is "the responsible administrative head of his school." He is the liason officer upon whom the impact of official and public requirements is made and from whom impulses emanate which actuate teachers and pupils.

Only in a few quarters is this ideal recognized in practice. Lip service may be done it but true homage is not paid. No effective restraint upon the tendency to multiply the clerical and administrative duties of the principal, and no consistent policy of securing principals whose training will permit them to measure up to the possibilities of their position are generally observable. For the system has produced in too many instances the type of principal which it has encouraged. Official displeasure is far more likely to be expressed because of a faulty report than because of failure to lead or to inspire. Thus the principal becomes a reporting officer. Since economy must be exercised, and since he will be praised for it, he distributes penholders and pencils with penurious care. He develops a system for doing this on the basis of the least good to the greatest number. With a concentrated prudence born of long experience, he doles out the supplies for which he is held so rigidly accountable. We know of one prin-

cipal—and a very successful one, too, as success is usually counted—who reduces all his supplies to the common unit of the value of one white 6 by 9 pad. A certain number of pads are allotted to each teacher per semester. She can take the pads either as such or in their equivalent of note books, rulers, or raffia. If by rigidly denying material to the children she accumulates credit in terms of pads, she might perhaps obtain a dozen scissors, or a set of paint boxes.

There is no real reason why the principal should perform such service. He should be relieved of it, not only for the sake of releasing his time for supervision (which is his proper service), but also because as a clerk he is a highly paid failure. Moreover, there should be an ideal in the minds of superintendents and board members in virtue of which principals trained for educational leadership should be selected. In particular, the principalship should not be merely the safe berth of a superannuated teacher. The principal's office is one that requires enthusiasm, qualities of leadership, plasticity, training and study. Exceptions occur; but it is doubtful if persons who for the most part have passed beyond the age of forty at the time when they begin their careers as principals are likely to render the type of service which an exalted conception of the principalship would require, and which under any circumstances the teachers and pupils should receive.

Thus is the time-honored and potentially powerful office of the building principal abased. Boards of education and superintendents could perform few more constructive acts of administration than to restore the principalship to its pristine dignity.

B. R. B.

Reviews and Abstracts

HAGGERTY, M. E. and others. *Virginia public schools; Part II, educational tests.* Yonkers-on-Hudson: World Book Company, 1921. xii + 235 pp.

Part II of the Virginia survey report deals with tests, educational and mental. It was prepared under the general direction of Dr. M. E. Haggerty of the University of Minnesota who is directly responsible for the chapters dealing with tests in arithmetic and general intelligence as well as certain general chapters concerning the work of the Division of Tests. Other chapters were prepared as follows: reading by Mrs. Cecile White Flemming of the Wisconsin State Department of Education, primary reading by Dr. M. E. Noonan of Harris Teachers' College, handwriting and spelling by Dr. F. S. Breed of the University of Chicago, high-school composition by Mr. Earl Hudelson of West Virginia University, and elementary algebra by Dr. H. G. Hotz of the College of William and Mary.

The Thorndike Reading Scale, Alpha 2, and the Haggerty Test, Sigma 1, were used in measuring ability in reading; tests in spelling were made up of words selected from the Ayres Scale; the Starch Scale was used in measuring handwriting; the Woody Scales and Exercise II of the Haggerty Intelligence Examination, Delta 2, were used in arithmetic, the former for fundamentals and the latter for reasoning; intelligence was measured by the Haggerty Tests, Delta 1 and Delta 2, and the Haggerty reading test, Sigma 1; the Nassau County Supplement to the Hillegas Scale was used in evaluating specimens in English composition; and Hotz's First Year Algebra Scales, Series A, Addition and Subtraction, Equation and Formula, were used in measuring the ability of first year high-school pupils in algebra. Although the list of tests used, as given in chapter I (p. 2), enumerates the Courtis tests, Series B, for arithmetic, and Exercise I of the Haggerty Intelligence Examination, Delta 2, for reading, no mention of the results secured through the use of these tests is made in the chapters dealing with those subjects.

In general, Virginia children fail to reach satisfactory standards. Although the best city white schools do as well as typical American city schools in reading and spelling, they fall behind in other subjects; while other city white schools, rural white schools, and city and rural colored schools lag far behind acceptable standards—in some cases as much as four years.

The book is a strong condemnation of the one-teacher school. "Almost one half of the white children (44 per cent) of Virginia receive their education in one-teacher and two-teacher schools. Such schools are the only educational provision for two-thirds of the colored children of the state. The product of schools of this type is, therefore, of tremendous significance in the educational life of Virginia, and because this product is of such a deficient character, the

small rural school constitutes one of the most serious educational problems, not only in Virginia, but throughout the South." (p. 194)

Recommendations are summarized as follows: "Chief among the necessary changes are: (a) the passage of an effective compulsory education law; (b) the lengthening of the school term to a one-hundred-eighty-day minimum; (c) improvement in the qualifications of teachers; (d) the increase of supervision, particularly of the rural schools; (e) a reduction of the one-room schools wherever possible in favor of consolidation; (f) the restriction of one-room schools to five grades; (g) improvement in the classification of children; (h) the organization of special classes for backward and superior children; (i) the employment of standard educational tests in measuring the progress of children and the efficiency of instruction; (j) the creation in the State Department of a bureau of educational investigation; (k) the creation of similar bureaus in all city and non-city divisions where conditions permit." (p. 13)

Particularly commendable is the plan for training public school officials and college and normal school teachers and students through participation in survey methods. According to this plan "About two hundred and fifty prospective teachers, fifty teachers now in service, and twenty-one professors in normal schools and colleges engaged from one week to two months each in the giving and scoring of tests. These remain in the school system to utilize and carry forward the work begun by the survey staff." (p. 6)

Naturally, a book written by a number of individuals will, unless very carefully edited, lack coherence and a definite plan for stating results. For example, the factor of over-ageness is very carefully considered in interpreting the results of reading tests in chapter III but is not mentioned in the arithmetic chapter and is merely alluded to, in making a comparison of race scores, in handwriting (p. 79). Again the chapter on arithmetic does not adapt standard scores to Virginia's seven-grade system as does the chapter on reading. And again graphic methods are considered necessary for the presentation of the results in intelligence, spelling, and handwriting; but only one graph helps the reader to interpret the results in reading, while the chapters on arithmetic, high-school composition, and algebra contain none. The reader notes the absence of measures of variability in some chapters but finds them carefully calculated in others.

There is considerable evidence of editorial haste or carelessness in proofreading. Table 72 provides blank spaces for average deviations but does not supply them. Some tables indicate by the use of heavy type or parentheses the intervals in which the medians lie, others do not. Figure numbers and data are rather frequently misquoted.

One easily finds the usual statistical errors. They occur in tables 17, 20, 21, 52, 62, 89, and elsewhere. Figure 11 is presumably drawn from the data of table 52 but is incorrect. Figures 14-20 should each contain 100 dots. They contain various numbers. None of the errors detected by the reviewer is very significant but their effect upon the reader is such as to make him skeptical of the accuracy of computations which he is unable to check.

The book, however, is a valuable contribution to our library of test literature. It should be a source of inspiration and suggestion to the school people of Virginia and of the nation.

R. L. MORTON

Ohio University

FROST, NORMAN. *A comparative study of achievement in country and town schools.* (Teachers College, Columbia University, Contributions to Education No. 111) New York: Teachers College, Columbia University, 1921. 70 pp.

This little book constitutes a really valuable addition to the literature of educational tests and measurements. Its aims as set forth by the author are: "First, to show that the difference in performance in school subjects of children of different ages can be obtained; second, to show that this difference is a measure of school efficiency which may be used to measure schools or school systems; third, to apply this measure to a system of country schools; fourth, to compare the results in this country system of schools with the results in certain city school systems."

The author starts out by summarizing the results of similar comparative studies of achievement in city and country schools. Then, in introducing his own study, he calls attention to the fact that all previous comparisons have been made on a basis of grade standards, which basis assumes equality of grades in the school systems compared. The author holds that such equality of grades is not general, particularly as between city schools and country schools. He avoids this difficulty by substituting age differences in achievement for grade differences as a basis of comparison. He thus obtains data readily comparable with results obtained in other systems, regardless of the diversity of conditions in the different systems, such as length of school year, teaching standards, curricula, amount of retardation, and the time of year at which the tests were given. This idea alone is worth many times the price of the book to thoughtful investigators.

The country schools tested for the purposes of this study were the 68 public "white" schools of Madison County, Kentucky. Most of them had a school year of only six months. The results are tabulated separately for the schools having a six-months term and for those having a nine-months term. Hence some interesting and significant comparisons are made between the achievements of the pupils in the two classes of schools over a period of three years.

Three tests were used, namely, Trabue Completion-Test Language Scale B, Courtis Standard Research Tests in Arithmetic Series B, and Thorndike Silent Reading Scale Alpha 2.

The results are compared with data obtained from surveys made in seven cities and large towns where the same tests had been used. All the data are set forth separately for each test in numerous well-arranged tables, together with summarizing tables which strikingly exhibit the relative efficiency of the several school systems as measured by the tests. The evidence as presented

is strongly in favor of the larger school systems and the longer school year. Hence, it is concluded that city and town schools are much superior in point of efficiency as measured by pupil achievement.

The truth of this conclusion depends largely upon the accuracy of the data with which comparisons were made and the validity of these tests as real measures of what the school accomplishes for the child. However, both these points are recognized and admitted by the author. He was presumably limited in his choice of tests by the necessity of using those which could be given to children of all grades, and by the necessity of using tests which had been rather widely used in other surveys in order to have data for comparative purposes.

The book is written in a very readable style. The author and his coworkers appear to have considered and tried to eliminate as far as possible every factor that would tend to lead to inaccurate and misleading conclusions. On the whole this study seems to represent careful and thorough work in attempting to add something worth while to what we *really know*.

S. S. Brooks

Winchester, New Hampshire

Branom, Mendel E. and Branom, Frank K. *The teaching of geography.* Boston: Ginn and Company, 1921. 292 pp.

Since the World War keen interest in the study of geography has been marked by new text books, new courses of study, and new books on methods of teaching. This is one of the latest books on the teaching of geography. It has been written by two men who have had wide experience in teaching geography, to normal school students.

The authors' point of view is indicated in the three following statements which appear in the preface: (1) "It is possible to organize the geography course of study in such a way that the dominant view point will be geographic and at the same time will permit of the interpretation of materials about life centers"; (2) "geography is primarily a study of the interlacing relations of life forms and nature. It is a study of place relations and the adjustments which man makes to the geographic regions with their varied resources"; (3) "An attempt is made to relate the work to the child's interests and experiences and to give the child a worthwhile motive, so that he will attack the problems whole-heartedly."

The modern view points and aims of geography teaching are clearly explained in the early chapters of the book. Later chapters suggest the criteria for a modern course of study, sources of materials for enriching the study, the organization of the class for study and recitation periods, and teaching pupils how to study geography. The authors explain and illustrate the different methods which may be successfully applied to the teaching of geography, such as: the problem method, the argumentative lesson, the topical outline, the journey method, type studies, the story, dramatization, and the project method. It is the authors' opinion that no one method should be exclusively used, but that methods should be varied to fit the type of material to be presented. The concluding chapter is devoted to measuring the results of geography teaching.

The few geography tests now available are discussed, and types of test exercises which might well be used are suggested.

The value of the book is greatly increased by suggestive questions for study and by bibliographies at the end of each chapter. Teachers of geography and normal school students will find this book an excellent aid both in furnishing practical suggestions for teaching and in presenting the modern view point of geography.

P. R. STEVENSON

Ohio State University

MITCHELL, LUCY SPRAGUE. *Here and now story book.* New York: E. P. Dutton and Co., 1921. xii + 360 pp.

Mrs. Mitchell has given us in this book new purposes in story telling, and suggestions for experimentation. The *Here and now story book* is neither a new collection of stories nor merely a collection of new stories. It is rather a full and well illustrated report of the author's experimentation in story telling and story composition with children from two to seven years of age.

The first half of the introduction is devoted to a discussion of the content of the stories, which are distinctive for being the expression of the experiences and interests of the children for whom they were told or written and who themselves assisted, either directly or indirectly, in the composition of many of the stories. The stories for the youngest children were composed and told while the child was meeting the experience, one purpose being to deepen the experience by "giving it some pleasurable expression" and another to afford the child an opportunity for creative expression. Mrs. Mitchell emphasizes the fact that she attempts "to follow honestly the leads which here and now the children themselves indicate in content and in form, no matter how difficult or strange the going for adult feet."

In the latter half of the introduction the form of the stories is very fully discussed. In criticising the popular custom of casting children's stories in a mould to the adult's liking, Mrs. Mitchell says, "The adult more often than not has presented his stories and verse to children in forms which the children could not like because they literally could not hear them!" It is suggested that we get the form or pattern for our stories from two sources, the children's own spontaneous art forms, and those story and verse patterns which make an almost universal appeal to children. A number of illustrations are given of the spontaneous compositions of children ranging in age from twenty-two months to six years. Emphasis is laid upon the need of subject repetition, and the use of refrains as sign posts by which the child can keep the continuity of the stanza or the story.

The reader may be shocked by the use of some expressions common among children but not heretofore accepted as permissible for the adult in telling stories to children. For example, the author defends this "subject repetition" as being used by the child in helping himself back to the subject and illustrates its use by the child's statement, "The little boy who lived in this house, he did so and so—" In her stories, however, Mrs. Mitchell sometimes uses

the pronoun immediately after the subject, thus separating a subject and predicate which might otherwise be in juxtaposition. We know that some children who never hear subject repetition never themselves use it, and also that there are other ways in which the mind may be taken back to the subject. Her argument for this is not convincing. Mrs. Mitchell does not say how far one should vary from the grammatically correct in telling stories to children who rarely hear a correct English sentence before they go to school. All the illustrations are taken from her own experience in the City and Country School.

Much stress is laid upon the importance of getting the children to create stories of their own, to play with words. The influence of their observations and expressions is observed in such passages as the following:

> "The lightning is all shiny."
> "In the summer-time the wind goes like breathing,
> But in winter storm it growls and roars."

Although these stories and poems were written particularly for the children of the City and Country School, a number of them deal with experiences and observations so common among children generally as to be equally suitable to thousands of children elsewhere and far from New York. Among the poems and rhymes are some gems that give promise of becoming childhood favorites. Throughout the stories there is much use of tone and color and the style is rhythmic, simple, and beautiful.

Altogether, Mrs. Mitchell has given us in the Here and Now Story Book something new to work upon. With suggestions for stories for the youngest children, and an enlarged field for story work in general, we may feel more free to discard some of the traditional stories which have been used only for want of something better. The art of story telling may become a finer art than ever before, demanding an observation and study of childhood experiences and expressions, a judgment as to suitability of material and form, an originality, and a skill with language which may bring forth a truly beautiful and appropriate literature for childhood — a literature adapted to varying types of children living in varied environments. The need for something more and something better than we have had should lead others to follow up Mrs. Mitchell's experiment in order that we may have the full benefit of whatever it has to offer.

MYRTLE L. KAUFMANN

Elementary Supervisor, Logansport, Indiana

Educational survey of the Philippi school system. Philippi, West Virginia: Board of Education, 1921. 39 pp. Distributed by Public School Publishing Co., Bloomington, Illinois.

This survey is not pretentious; yet in a simple, direct manner it describes the educational needs of Philippi, in such a way that anyone, even though uninitiated into the mysteries of educational tests, can understand the situation and get helpful suggestions for remedial work. It was written primarily for the school patron and the teacher who has had little experience with

measurement work. As might therefore be expected, there is little help afforded to those who have been using educational measurements for some time, except it be in the conciseness of presentation.

After a brief statement of the industrial and historical features of the city, in which considerable pride is manifested, the school building is described. The contrast is striking. The building no longer adequately cares for the legitimate educational activities and its poor condition is thought to account for the abnormal amount of sickness. In three chapters, are presented "The Classification of Pupils," the "Results of Education Tests," and the "Superintendent's Records" and summary. A comparison of a chronological age-grade table with a mental age-grade table shows that the excessive retardation is not justified by the intelligence level of the pupils and that there is much need for a reclassification. The following educational tests were given: Courtis' Standard Research Tests in Arithmetic, Series B; Monroe's Standardized Silent Reading Test; Thorndike's Visual Vocabulary Test; Cavins' Test in American Poetry; Hudelson's English Composition Scale; Charters' Diagnostic Language and Grammar Test (verbs and pronouns); Ayres' Handwriting Scale (Gettysburg Edition); Buckingham's Extension of the Ayres Scale for Spelling; Harlan's American History Test; Courtis' Supervisory Tests in Geography; Rugg and Clark's First Year Algebra Test. Only in the more formal phases of English work was the school up to standard. In the discussion of each test are very practical suggestions for remedial measures.

This survey was carried on by the Department of Education of West Virginia University under the direction of Professor L. V. Cavins. It is typical of the assistance the university is ready to render to schools of that state.

W. W. Coxe

Ohio State University

Wagner, Charles A. *Common sense in school supervision.* Milwaukee: The Bruce Publishing Company, 1921. 201 pp.

"The art of supervision is quite young and the function is not yet sufficiently differentiated to have yielded any large body of accepted facts or to have developed any long series of settled procedures." (p. 187). "Wrought out of experience and wrought into print, they (the chapters) represent just what one supervisor has found workable." (p. 179). "The most that is claimed for those chapters is that they do constitute an attempt at an answer (to the problems of supervision), though they do not assume that it is the only answer nor that it is the best answer." (p. 179). "If speaking from experience be regarded a valid reason for speaking at all on a subject, then these chapters have the fullest possible warrant; thirty years of practicing, of speaking, and of writing the ideas presented." (p. 3).

The book is just what its title indicates — a frank, practical exposition of the subject by one who has had much experience.

The various chapters overlap to a considerable extent, which is indicative of a lack of clear analysis of the subject. The reader is occasionally startled

by the use of an unusual word, as, for example, *frequence*, which is used nine times on page 21. Some sentences give one a positive shock. The following illustrate the cause of the shock:

"Nothing but supervision can secure unity, use of the same textbook does not secure it, use of the same course of study can not assure it." (p. 157).

"Your careful note taking and your serious effort to help yourself, will impress upon him (the supervisor) that you are both careful, studious and seriously in earnest." (p. 170).

The author's general point of view is that supervision exists because teachers are unprepared and to meet "the demand for unity, continuity, sequence of instruction and for equivalence of opportunity for all the children of a school system." (p. 8). The better preparation of teachers for public school service seems to the author quite impossible. "The normal schools can not possibly find places and classes enough for the students to get any real training in the normal school." (p. 188). Therefore, the only way to an increased effectiveness of public school work is through more and better supervision.

The great objection to supervision is the personal, as opposed to the professional, point of view. Just how this objection is to be removed, so long as teacher and supervisor both remain human, is not convincingly set forth.

Many persons now engaged in supervision and without adequate preparation for their present responsibilities will read the book with interest and profit.

JOHN A. H. KEITH

State Normal School, Indiana, Pennsylvania

News Items and Communications

This department will contain news items regarding research workers and their activities. It will also serve as a clearing house for more formal communications on similar topics, preferably of not more than five hundred words. These communications will be printed over the signatures of the authors. Address all correspondence concerning this department to Doctor E. J. Ashbaugh, Ohio State University, Columbus, Ohio.

A circular letter sent to the school people over the country asking for communications has been receiving a splendid response. If you, reader, either did not receive or have not answered such a communication, please feel it your responsibility to send such material to the department editor.

Superintendent Ira J. Bright, Leavenworth, Kansas, reports progress on the one-half million dollar building program. The bond election carried approximately two to one in spite of business depression and increased tax rates. During the short campaign publicity in newspapers, movies, printed pamphlets, window cards, wind shield pasters, three minute speakers, and school parades were all used. The slogan throughout was "True Americans will vote for modern schools."

Superintendent F. W. Johansen, Cherokee, Iowa, resigned January 30 to become Vice-President of the First National Bank of that place. It is a distinct loss to have school men who have made the record of Mr. Johansen leave the service.

We have recently received a copy of a Physics test devised by H. W. Tilton, University of Pittsburgh. It is of the sentence-completion type and anyone interested in this phase of the measurement work should write to Mr. Tilton.

Superintendent W. H. Richardson, Cuyahoga Falls, Ohio, reports that they have been using the "work-study-play" plan of organization for six years and have found it entirely satisfactory. A new high school building is nearing completion, the controlling principle in the construction of which was "complete use of every part of the building and a duplication of uses of rooms." The building will cost $300,000 and accommodate 600 pupils.

Superintendent Theodore Samm, Council Bluffs, Iowa, reports that the first year of their School Savings Department showed a total deposit of $76,986 with 4,477 children depositing. That 4,063 children had active accounts amounting

to $26,008.22 on December 31 shows that the thrift spirit in the schools survived through the Christmas shopping period. We shall be glad to receive similar information from other communities which have school savings.

We are just in receipt of a copy of "The Idaho Teacher" and note that the contents indicate that the magazine lives up to its statement of aim, "A journal devoted to the interests of education in Idaho." All phases of education from kindergarten to university are reported with live news. The report of the State Association's Committee on teacher shortage and salary shows 10.8 percent new to the profession, 55.7 percent new to the position, and only 5 percent holding present position more than three years. It is interesting to note, however, that the supply practically meets the demand and that salaries have remained the same as last year.

The superintendent of public instruction of Oregon has sent us their Arbor Day Manual for 1922. Aside from material suitable for Arbor Day exercises, it contains a large number of records of native shrubs with descriptive material which will go far in helping the children to recognize the flora of their own state. This is especially commendatory.

Superintendent J. O. Engleman, Joliet, Illinois, was elected president of the Illinois State Teachers Association at its holiday meeting. For the past three years he had been chairman of its legislative committee.

A committee of five, three principals and two teachers, elected by the teaching staff is working, by invitation of the Board, with Superintendent Engleman and a committee of the Board upon a revision of the Board's Manual. The step was taken in order that the teachers might have a voice in the rules, especially those that govern salaries, sick leave, teachers meetings, and other matters concerning the teachers themselves. Every teacher on the staff is working this year upon a revision of some phase of the course of study.

Professor R. A. Kent, Dean of the School of Education, University of Kansas, writes that plans are being formulated for a conference at the university "of school superintendents and others who may be interested in the discussion of the problems connected with educational research particularly in the public school field." Conferences of this sort are becoming surprisingly general and the extent to which they are attended indicates their popularity.

Richard D. Allen, Director of Vocational Guidance for the Public Schools of Providence, Rhode Island, points to the fact that the program of vocational guidance in Providence manifests a widely different "slant" from that which work of this sort usually takes. He does not believe in combining the work of vocational guidance with the duties of the attendance officer, the school census officer or the director of vocational education. He rightly points out that when so combined the guidance work tends to become supplementary or incidental to other activities. It is Mr. Allen's idea that the whole matter of

vocational counciling should be carefully surveyed from the point of view of the *educational system* rather than from that of juvenile employment. He proposes such a survey as a research problem.

We have received a new form of report card from Superintendent Richard C. Tobin, of Ansonia, Conn.

This card is a small four page folder which provides for five ratings during the year and for a summary at the close of the year.

The distinctive feature of the card is the fact that not only in respect to scholarship but also in respect to certain other qualities such as attitude toward school work, character of the pupil's response in class, and his conduct are rated by each teacher separately. The arrangement by which this is accomplished is decidedly ingenious. We think that it will be of some interest to the parents of pupils in the Ansonia schools to note the extent to which their children show differences in attitude, conduct, etc., as they come in contact with different teachers.

Regional Conferences on Educational Measurement

There are 207 district superintendents of schools in New York State. These superintendents have general supervision of all schools in rural communities and in villages under 4500 population.

In the spring of 1921 the State Education Department arranged a series of regional conferences for district superintendents on the general subject of classroom supervision. At each of these conferences one hour was given to the discussion of the supervisory uses that might be made of standard educational and group intelligence tests, and it was suggested that the Department would be willing to send a representative to meet with superintendents to discuss the question at greater length. Following these conferences the Department received so many requests for assistance that it was impossible to meet them all with the help at hand. In early September, a circular letter was sent to all district superintendents stating the difficulty that the department had met and suggesting that a number of regional conferences might be arranged at the several normal schools of the state, if district superintendents so desired.

Within a week replies had been received from 120 superintendents asking that such conferences be arranged and stating the subjects in which they would be most interested to study tests. Conferences were arranged at eight centers; Oneonta, Buffalo, Brockport, New Paltz, Cortland, Potsdam, Plattsburg, and Albany. The first of these conferences was held October 10 and 11, the last December 8 and 9. The program was arranged specifically, so far as possible, to meet the needs of district superintendents. In the correspondence practically all of the superintendents had signified a desire to study tests in reading and arithmetic; and a large majority desired group intelligence and spelling tests.

Each conference lasted two days. The first morning session was primarily devoted to a discussion of the criteria to be used in selecting tests for supervisory purposes; and to studying carefully the technique and directions for giving certain of the tests to be demonstrated during the conference. The

afternoon session was given to demonstrating four or five of the tests selected or agreed upon at the morning session. In each conference the Monroe Silent Reading Test Revised and either the Burgess Silent Reading Test or the Thorndike-McCall Reading Test were demonstarted. The arithmetic tests considered were the Courtis Research Tests, Series B; the Courtis Standard Practice Tests; Woody Arithmetic, Series A; and Woody-McCall Mixed Fundamentals. Two or more of these tests or scales were demonstrated at each conference. The usual form of procedure was for the specialist in educational measurements from the department to demonstrate the giving of the National Intelligence Test, Scale A, and the Monroe Standardized Silent Reading Test Revised to the same pupils. This demonstration was usually followed by a superintendent (usually one who had had previous experience) demonstrating the giving of two of the reading tests,—the Monroe and the Thorndike-McCall, or Burgess Scale to a second class. Next, two or more of the arithmetic scales were given to different classes by superintendents. In each case the choice of scales had been determined by the majority of the group in the morning discussion. The demonstrations were invariably followed by a spirited conference on the technique of giving tests. Usually enough errors had been made to make this discussion quite effective. When time allowed, one set of papers was scored in the afternoon session.

The second morning was given to scoring papers, tabulating results, and interpreting data. Points that aroused unusual interest in the discussion were the relation between rate and comprehension scores in reading; between reading scores and group intelligence test scores; and the means of using the arithmetic tests for diagnosing the needs of classes and individual pupils. In most of the conferences the classroom teachers or critic supervisors joined in these discussions so that superintendents had the opportunity to compare the tests results with the teacher's knowledge of the individual pupils. Throughout these morning discussions there entered from time to time questions and discussions of what the superintendent might do with similar data from his own schools toward improving classroom instruction.

The second afternoon session was usually not more than two hours in length and was devoted to an informal round table discussion of educational measurements.

It was the aim of these conferences to give superintendents such specific, concrete acquaintance with the giving, scoring, tabulating, and interpreting of two or three tests that they might, with a considerable degree of confidence go home and use these tests in their own supervisory work. The dozens of letters already received by the department describing what superintendents are doing, or asking for further specific information indicates that a considerable majority of them have already undertaken some measurement work.

One feature seemed significant in the conferences. During preceding years between 20 and 30 district superintendents had used one or more standard educational tests. With only one exception these twenty-odd superintendents attended the conferences and took a leading part in the discussion.

Conferences were attended by 158 district superintendents, three city or village superintendents, 23 village and elementary principals, and 10 super-

visors or teachers—a total of 194. These figures do not include those people who came in for only one or two sessions, or the many normal school teachers and supervisors who attended part or all of the conferences. Five cities and villages which were represented in the conferences either directly by the superintendent or some one representing him have already introduced a part of the measurement program in their work or have indicated their intention of doing so.

It is estimated that reading and arithmetic tests demonstrated in the conferences will be used as a basis for improvement of instruction, during the current year, in not less than 5000 classrooms of the State.

J. CAYCE MORRISON

State Department of Education, New York

Rhode Island and Massachusetts

The Journal of Educational Research for October very courteously published my comment upon "The Education of Teachers" and a reply by Doctor W. Randolph Burgess, in which he says, in defense of his position: "This statistical method has always been open to the criticism of the person who calls attention to specific cases which the method does not measure." May I venture to suggest that a statistical method that will not stand the test of correlation to fact bears unmistakable indication of a want of validity. I did not, however, criticise the statistical method generally, nor the particular statistical method used by Doctor Burgess. I did criticise his application of the method and his failure to use facts when facts were available; and I pointed out that the application of facts that are unquestioned would change the results of his study as to index numbers and relative standings for the States. In this connection may I not be permitted to say that Commissioner Payson Smith of Massachusetts in a public address has accepted the validity of the corrections that I suggested to Doctor Burgess, and has recognized Rhode Island's leadership in preparing teachers for service and in the preparation-personnel of the teaching staff. His predecessor, Commissioner David Snedden, did likewise in his report for 1914, in which he wrote: "Except Rhode Island, no state surpasses Massachusetts in provision for the training of teachers." Were I in Doctor Burgess' position, I should rejoice that the correction of a single fact item would square my statistical theory with fact. There is certainly a genuine comfort in proving the statistical method valid and accurate, by applying to it the test of experience and fact and finding correlation. Indeed, when statistical theory and fact do not coincide, I question the validity of the theory.

With reference to Doctor Burgess' second article, published in the October Journal, "The Rate of Progress in Teacher Preparation," it is worthy of comment that he did not use for Rhode Island in Table 1 a corrected index figure warranted by the fact that in Rhode Island graduation from normal school has meant 2½ years of professional training beyond graduation from high school. The index number for Rhode Island should be 2.39 instead of 2.00, as Doctor Burgess calculates it. Rhode Island should have first place, instead of second.

Similar index figures for Rhode Island for eleven years from 1910 to 1920, are interesting as showing steady advance and also as indicating the probable path of progress. These numbers and the facts on which they are calculated follow:

Year	College Graduates	Normal School Graduates	Number of Different Teachers	Index Number
1910	341	1471	2371	2.13
1911	348	1510	2394	2.15
1912	363	1563	2428	2.21
1913	368	1620	2469	2.24
1914	396	1783	2586	2.34
1915	386	1839	2680	2.30
1916	412	1939	2773	2.34
1917	428	1998	2844	2.36
1918	441	2055	2934	2.35
1919	422	2161	2966	2.39
1920	422	2164	2971	2.39

These index numbers if used in Table I as printed in the October Burgess article would establish leadership for Rhode Island in every year from 1910 to 1920, and also indicate in a general way that Rhode Island in 1910 had a higher index number than its nearest rival had in 1920; or that Rhode Island is approximately ten years ahead of any other state in the personnel of the teaching profession.

In Doctor Burgess' Table I, the index numbers for Massachusetts indicate in a general way the progress that a state system may make in ten years, starting with an index number of 1.65. These Massachusetts index figures are strongly corroborated by a similar group of index numbers for Rhode Island in the period from 1899 to 1909. Placed side by side in a table the correlation is obvious and striking:

Rhode Island		Massachusetts	
Year	Index Number	Year	Index Number
1899	1.66	1910	1.65
1900	1.71	1911	1.65
1901	1.74	1912	1.68
1902	1.80	1913	1.68
1903	1.87	1914	1.75
1904	1.92	1915	1.82
1905	1.92	1916	1.90
1906	2.03	1917	1.98
1907	2.06	1918	2.05
1908	2.04	1919	2.05
1909	2.10	1920	2.05

New Hampshire, starting with 1.71 in 1920, may reasonably be expected to advance to 2.10 by 1930, if it follows the general experience of Rhode Island and Massachusetts; perhaps from 2.10 to 2.40 by 1940, if the progress made in Rhode Island is duplicated; and from 2.40 to 2.60 by 1950, assuming that the

apparent law of diminishing returns operates as indicated in the experience already accumulated. The three states should stand as indicated in the following table at 10 year intervals from 1910 to 1950:

	1910	1920	1930	1940	1950
Rhode Island	2.10	2.40	2.60	2.70	2.80
Massachusetts	1.65	2.05	2.35	2.55	2.65
New Hampshire	1.00	1.70	2.10	2.40	2.60

Massachusetts and New Hampshire should gradually approach Rhode Island, which will continue to maintain leadership because of a long advance, though rising more and more slowly, as will be true also of her two closest rivals. On these statistics and facts I venture to say that Doctor Burgess is in error in his estimate that Rhode Island will be in fifth place by 1950.

Finally, a few facts may indicate the course of future events. New Hampshire has made splendid progress in recent years doubtless largely because of the operation of a law providing state support for schools taught by normal school graduates. Massachusetts is at present becalmed, awaiting a fresh impulse. Rhode Island has reorganized its normal school as a college of education. An immediate effect thereof has been to double the entrance class; and the indications are that within five years enough trained teachers can be graduated from the 2½-year course, as that is continued, to replace all teachers with less than normal school or college graduation. That means at least an index number in excess of 2.50. Moreover, the college in June will graduate a first class of teachers with four years of training, and will continue to do that year after year. Ultimately the 2½-year course will disappear and normal-school graduation in Rhode Island will mean four years preparation. Still more, 1100 teachers in service, mostly normal school graduates, are comprising more than 36 percent of all the teachers in the state, are enrolled for summer school and extension courses, leading to the completion of four years of preparation for teachers. In other words, Rhode Island has already undertaken measures that will produce an index number 4.00. May all the states join in the race; for that will mean an advance in teaching that will produce inestimable improvement in our public schools.

CHARLES CARROLL.

Rhode Island College of Education

Handicapped School Children and How to Meet Their Needs in Ohio

One of the most significant contributions to the science of education made during the last quarter century—especially the last decade—is the demonstration of the large individual differences which exist in the health, physique, intelligence, and emotional, social, moral, and pedagogical development of school children. Many factors are responsible for the variations which represent inferiorities and disabilities, e.g., the biological tendency toward variation, largely ineradicable, neglect, diseases, defects, and defective conditions in the ancestry and in the environment. Among the most important types of handicapped children needing special treatment are the blind, semi-sighted, deaf,

semi-deaf, anemic, crippled, feeble-minded, backward, delinquent, speech defective, and word-blind. These groups of maladjusted children create the most difficult problems which the schools and society have to face. The responsibility for the improvement or rehabilitation of these children devolves far more upon the schools than upon any other agency. How shall the schools discharge this responsibility?

Not by ignoring the individual differences and treating the children as normal, as has been done from time immemorial, but by discovering, understanding, and making adequate provisions for all the significant types of variations, either by furnishing special aid in the regular program of studies or by supplying differentiated instruction and training in different kinds of special classes. In order adequately and economically to organize a system of special classes in Ohio it is necessary to carry out a state-wide program which shall include: (1) The establishment of various types of special classes to meet the needs of the handicapped children found in all sections of the state. (2) The manning of the classes by teachers especially trained for the type of specialized work in which they are engaged. Ohio has made special provisions for the training of teachers of handicapped children by establishing (in February, 1920) a Bureau of Special Education under the administration of Miami University, which will be located in Dayton in June where the public schools will afford adequate clinical facilities, and a sufficient number of different types of special classes for observation and practice teaching. (3) The adequate examination and certification of the pupils before assignment, by local and state examining bureaus. The state has already established the Bureau of Juvenile Research, charged especially with the examination of delinquents committed to the guardianship of the department of public welfare, or any minor received by this department for observation, and the Bureau of Special Education, which aims to advise with public school authorities regarding the organization of special classes and to test and examine school children, while many public school systems and higher institutions of learning also support departments of tests and measurements. (4) The enactment of further legislation in the interest of handicapped school children. Ohio now grants exceedingly generous "state aid" for the support of classes for the blind, deaf, and crippled. We need a law, mandatory in character, providing for the establishment, under state subsidy, of special classes for the feeble-minded, for the delinquent, for the backward, and for the organization of work for the correction of speech disorders.

Various facts, gathered from a study of the results of special class work in a large school system, were presented in proof of the statement that differentiated and corrective training for maladjusted children is eminently worth while from the standpoint of the social returns.

J. E. W. Wallin

Miami University

Methods of College Teaching

It is said and seldom gainsaid that college teaching is of poor quality. Yet it is certain that some college teachers are more successful instructors than others. From this it is but a step to the conclusion that the successful have something to teach the less successful and the inexperienced. If, in each institution of higher learning, some way could be found to capitalize the ability of the best teachers and to make their methods better understood throughout the institution, it is unquestionably true that the institution would better fulfill the purposes for which it was created.

Yet whoever heard of conscious attempts in colleges and universities to improve instruction? Supervision is non-existant, and for one instructor to visit the classes of another is almost unheard of. Under these circumstances, the work of Dr. W. W. Charters at Carnegie Institute of Technology—a work which has for its purpose the improving of instruction in that institution—is worthy of attention.

From a report which reached our hands some time ago, it appears that Dr. Charters conducted a seminar in Methods of College Teaching. The first work of this group of students appears to have been a sort of job analysis of college teaching. In the mimeographed report of the work of the seminar, eleven items are recognized in this job analysis—such items as *learning to know students, apportioning time to material, methods of daily preparation, testing and grading, the assignments,* etc.

The plan was to select from the faculty of the institution a number of the most successful instructors and to find out from each of them what provisions they made in their teaching for the items which the job analysis had revealed.

The method by which the information was obtained from these successful instructors was in itself a matter of no small importance. It would have been easy to send each one of them a questionnaire, and this undoubtedly would have been the easiest method to adopt. Undoubtedly too, it would have yielded disappointing results. The method actually followed was to provide for a personal interview of each of the instructors who had been selected as having something to contribute in the way of successful methods of teaching. Moreover, the interviewers were themselves trained, each being provided with a standard form and instructed in its use.

The resulting report is worthy of a larger circulation than it has received. Many a teacher, even though he is already highly successful, would benefit by knowing the ways in which others who are undoubtedly successful accomplish their results.

Incidentally one gathers from reading the report the impression that the college instructor as represented by these successful individuals, is surprisingly conscientious and painstaking. The careful planning of work, the time consuming methods of recording and filing data, the elaborate preparation, the self-analysis, and the self-denial with which these men and women approach and pursue their task shows that least in these instances college teaching is both more exacting and more excellent than its undiscriminating critics have supposed.

B. R. B.

The Place of Final Examinations

The reforms in secondary education brought about during the last decade have shown a marked tendency towards more liberal practices and away from all formalism that smacked of academic pedantry. The results thus far are evidenced in broader curricula, enriched subject matter, and methods of instruction that aim to give greater vitality to the courses of study. When it was recognized that the high school bore as important a relationship to the community as to the college or university and when the colleges granted most liberal terms for admission, the formal final examination passed out of existence in many schools as a useless appendage.

On the surface the final examinations in high school were valuable only so long as pupils were required to face the ordeal of entrance examinations for college. Since they gave no additional evidence of scholarship, they were of no assistance to the teacher in determining the passing grade; and they did undoubtedly subject some pupils to a terrible nervous strain. Besides they were costly in time and energy, and they could hardly be justified as an artificial stimulus to scholarship or a worthy rod to prod the indolent. So in many high schools final examinations were discontinued.

Not all schools so acting, however, have been altogether pleased with the results of this liberalizing reform. Some have thought that there was a deterioration in scholarship, a let-down from serious purposes of study, and a loss in the tone of the school. Administrative problems have increased rather than diminished. A thorough study of the place of final examinations has in some cases resulted in a return to the earlier practice.

Most of the arguments against final examinations are unanswerable. As applied to the worst type of examinations they are sufficient to discredit them entirely. But the important place of examinations in the learning process is usually ignored, and it is this psychological basis for examinations that establishes their right to existence and furnishes the explanation for the school conditions resulting from their discontinuance.

Doctor S. S. Colvin, after discussing at length the various laws of memory, says, "In the elementary and secondary schools examinations, not as tests, but as incentives for keeping in mind the work that has been pursued over an extended period, should be emphasized more than they are at present."[1] Doctor W. C. Bagley, writing in criticism of the project method, expresses the same psychological principal from a different angle: "There is some indirect evidence that information gained primarily to solve an immediate problem is *not* so long retained or so easily recalled as is information that is mastered with the intent to make its mastery permanent. I have abundant experiences of my own that confirm this conclusion. . . . In fact the validity of the principle of 'learning with intent to remember' has been fairly well established by experimental evidence. . . . Psychologically, learning with the intent to master permanently comes very close to learning for its own sake."[2]

---

[1] Colvin, Stephen S. *The learning process*. New York: The Macmillan Company, 1913. p. 176.

[2] Bagley, William C. "Projects and purposes in teaching and in learning." *Teachers College Record*, 22: 289-90, September, 1921.

The results of an inquiry into the practice of a number of large high schools are given herewith. The evidence rather favors the formal final examination. The sins of malpractice seem to be well known but without terror to the alert supervisor.

From 48 high schools, 44 replies were received; 25 of which indicated that final examinations were held, the remaining 19 indicating that they were not held. The schools that give final examinations report positively that in 22 of the 25 instances the system works in general to the satisfaction of teachers and patrons. The remaining three schools express doubt on this point, but none of the 25 schools is positive that the system is not working well.

Examinations are prepared about equally often by teachers and by committees. Most commonly two teaching periods are devoted to the examination, and the results count from one-fifth to one-third toward the semester's rating. Far the greater part (17 out of 25) of the schools which use final examinations do not exempt from examinations students of superior ability. None of these schools conducts examinations unders the honor system.

Among the arguments given for final examinations are the following: (1) they prepare for college examinations and for competitive examinations of all kinds; (2) they develop ability to meet crises, to summon the best that is in one; (3) they lead to wise review and better organization of subject matter; (4) they secure better application and concentration on the part of teachers; (5) they tend to hold pupils to the end of the year; (6) they tell teachers where the weak points are; (7) they afford an incentive for pupils to make up work lost during absence or neglect; (8) they give a working standard that changes only gradually from year to year and is independent of the personality of any one teacher.

It is perhaps significant that the principals of the schools which do not give final examinations are not as certain of the success of their plan as are the principals of schools which have the examinations. Seventeen of the schools which do not give final examinations answered the question as to whether their system was working generally to the satisfaction of teachers and parents as follows: yes, 8; no, 3; doubtful, 3; principals personally preferring them, 3.

From the schools which do not give final examinations the following were listed as the best arguments for their policy: (1) cramming is not encouraged; (2) nervous strain is avoided; (3) the teacher can judge the ability of the student without the final examination; (4) such examinations constitute an unwise expenditure of time; and (5) they are unfair as a sole basis for promotion.

E. J. EATON

Principal South High School, Youngstown, Ohio